

Multispectral YOLO: Generic Feature Fusion Framework for Solar Active Region Detection

António Santos

up202008004@up.pt

LIACC, Faculdade de Engenharia, Universidade do Porto

Filipa S. Barros

LIACC, Faculdade de Engenharia, Universidade do Porto

J. J. G. Lima

Instituto de Astrofísica e Ciências do Espaço, CAUP

Rui F. Pinto

Institut de Recherche and Astrophysique et Planétologie, OMP/CNRS, CNES, University of Toulouse

André Restivo

LIACC, Faculdade de Engenharia, Universidade do Porto

Luís F. Teixeira

Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC)

Research Article

Keywords: YOLO, Multispectral, Active Regions, Object Detection

Posted Date: September 19th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7600800/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Machine Vision and Applications on April 9th, 2026. See the published version at <https://doi.org/10.1007/s00138-026-01807-y>.

Multispectral YOLO: Generic Feature Fusion Framework for Solar Active Region Detection

António Santos (up202008004@up.pt)^{1*}, Filipa S. Barros^{1,2,3†}, J. J. G. Lima^{2,4†},
Rui F. Pinto^{3,5†}, André Restivo^{1†}, Luís F. Teixeira^{6,7†}

^{1*}LIACC, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias,
Porto, Portugal.

²Instituto de Astrofísica e Ciências do Espaço, CAUP, Rua das Estrelas, Porto, Portugal.

³Institut de Recherche and Astrophysique et Planétologie, OMP/CNRS, CNES,
University of Toulouse, Avenue Colonel Roche 9, Toulouse, France.

⁴Departamento de Física e Astronomia, FCUP, Rua do Campo Alegre 687, Porto,
Portugal.

⁵Infor'Marty, Avenue Colonel Roche 9, Toulouse, France.

⁶Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC
TEC)Rua Dr. Roberto Frias, Porto, Portugal.

⁷Faculdade de Engenharia da Universidade do Porto (FEUP), Rua Dr. Roberto Frias,
Porto, Portugal.

[†]These authors contributed equally to this work.

Abstract

Monitoring solar phenomena, such as sunspots and active regions, is crucial for ensuring astronaut safety, telecommunications reliability, and predicting terrestrial events like auroras. Traditional methods for detecting these phenomena have limitations in accuracy and baseline maintenance. This paper presents a novel deep learning object detection method that leverages multispectral image data from satellites to enhance the detection of "sunspots" and active regions. Utilizing images from the SDO satellite and annotations from the DeepSDO dataset, we constructed a new dataset composed of aligned observations from HMI Ic, AIA 211 Å, and AIA 335 Å. We adapted and developed a stock YOLOv5-based model capable of handling and fusing any number of input images. Two fusion methodologies, early and late fusion, and three different fusion modules — CatFuse (simple concatenation), CBAMC (CBAM-based module), and TransEnc (transformer encoder) — were implemented and tested. Our critical evaluation of the models, supported by statistical analysis, proved the developed models to be statistically significantly different among themselves at a p-value of 0.05, and helped us to identify the best-performing model: CatFuse with early fusion, which achieved a mAP@0.5:0.95 of 0.52 and a mAP@0.5 of 0.94. This result was marginally better than the best baseline (YOLOv5 with a single HMI image) and comparable to other state-of-the-art models, demonstrating a modest but consistent improvement of multispectral image fusion for this task.

Keywords: YOLO, Multispectral, Active Regions, Object Detection

1 Introduction

Monitoring and understanding solar activity, particularly sunspots and active regions, has become increasingly important with the growing presence of human infrastructure and exploration beyond Earth. These features are critical for astronaut safety, the stability of telecommunications systems, and the forecasting of Earth-bound phenomena such as auroras. Sunspots, which manifest as dark regions on the solar surface due to intense, localized magnetic fields, are closely associated with dynamic solar events—including solar flares, radiation bursts, energetic particles, and coronal mass ejections (CMEs), that pose risks to astronauts and satellites, and can significantly disrupt Earth’s geomagnetic environment [1, 2].

Detecting active regions on the Sun has evolved from manual observations and basic image processing to advanced deep learning (DL) techniques that have transformed computer vision [3]. Despite these advances, most prior DL and machine learning approaches for sunspot detection have relied on single image modalities, primarily magnetograms or HMI (Helioseismic and Magnetic Imager) intensitygrams, although contemporary instruments capture multispectral data across diverse electromagnetic bands [4–9]. This narrow focus limits the exploitation of complementary information inherent in multispectral observations, potentially constraining detection accuracy and robustness.

To address this gap, we developed a flexible and extensible object detection model based on YOLOv5 [10]. YOLOv5 was chosen for its lightweight architecture, strong performance, ease of customization through YAML configuration files, and active community support [11]. Furthermore, concerns regarding constraints, the hardware limitations of the machine used, informed our decision to keep the development to this particular YOLO version, seeing as, naturally, multiplying the number of backbones increases the model size. Notably, existing YOLO implementations lack standardized mechanisms to process multiple image inputs, a limitation that our approach explicitly overcomes [12]. Our methodology involved several key steps:

Dataset Creation We meticulously constructed a large-scale, personalized multispectral dataset for sunspot detection by leveraging existing sunspot annotations from the DeepSDO dataset. This was supplemented with multispectral images (HMI Ic, AIA 211 Å, and AIA 335 Å) from the SDO satellite accessed via the JSOC database. The resulting dataset ensures consistency in both image source and annotation alignment [13].

Model Adaptation and Fusion Methodologies We adapted the YOLOv5 architecture to accept and fuse multiple image inputs. Two primary fusion strategies were implemented and rigorously evaluated: early fusion, where images are combined before the network backbone, and late fusion, where features extracted by separate backbones are merged later in the network.

Feature Fusion Modules To investigate different feature aggregation approaches, we developed and tested three fusion modules: **CatFuse**, a straightforward concatenation-based method; **CBAMC**, an adaptation of the Convolutional Block Attention Module (CBAM) that enhances relevant features via channel and spatial attention; and **TransEnc**, a Transformer encoder-based fusion module inspired by advanced attention mechanisms.

Comprehensive Evaluation We performed thorough testing and critical evaluation of all models using statistical analysis methods, including the Friedman test [14] and Wilcoxon post-hoc test [15], to assess performance differences and the impact of various fusion techniques and image modalities.

Our work demonstrates that early fusion strategies, particularly when combined with the CatFuse and CBAMC modules, outperform the best single-image baseline (HMI), achieving a mAP@0.5:0.95 of 0.52 and a mAP@0.5 of 0.94. In contrast, late fusion approaches proved less effective. The most sophisticated fusion module, TransEnc, failed to produce anticipated improvements in early fusion and was computationally prohibitive when applied in late fusion. Overall, our results are comparable to other state-of-the-art models in sunspot detection. This study establishes a robust foundation for multispectral object detection in solar physics by addressing the lack of generic, scalable implementations and reframing sunspot detection as a multispectral task that fully exploits the potential of available data.

2 Methodology and Data

An incremental experimental approach was adopted, progressing through dataset creation, model adaptation, and the development of fusion mechanisms.

2.1 Data Creation

A personalized multispectral dataset was constructed, as no readily available dataset met the required criteria, namely, multiple wavelength images of the same solar event with consistent annotations.

The DeepSDO dataset [13], containing 4,383 full-disk HMI Intensitygrams accompanied by astrophysicist-validated sunspot annotations, served as the foundation for our dataset.

A Python script was developed to fetch corresponding images from the JSOC database using SunPy’s drms package [16], [17]. The script fetched HMI Ic, AIA 211 Å, and AIA 335 Å images, as these are SDO images where sunspots are clearly visible. The original PASCAL/VOC [18] annotations were converted to the YOLO format.

A crucial processing step was added to crop AIA images by 35 pixels on all sides, via iterative fine-tuning and manual validation. This addressed a scale difference where the Sun appeared smaller in AIA images compared to HMI, ensuring annotations remained aligned across different wavelengths.

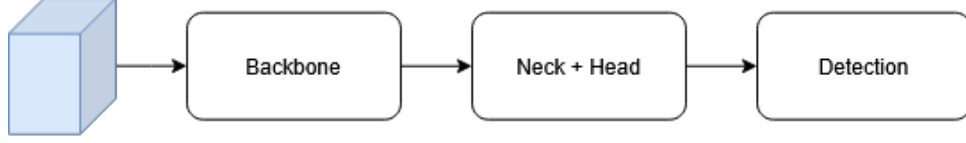
From the original 4,383 sunspot observations in DeepSDO, 3,397 multispectral instances were retrieved for training, and 861 from the test set, which we used as our validation set, with missing instances excluded due to incomplete image data. The final datasets are publicly available¹.

2.2 Model Adaptation and Fusion Modules

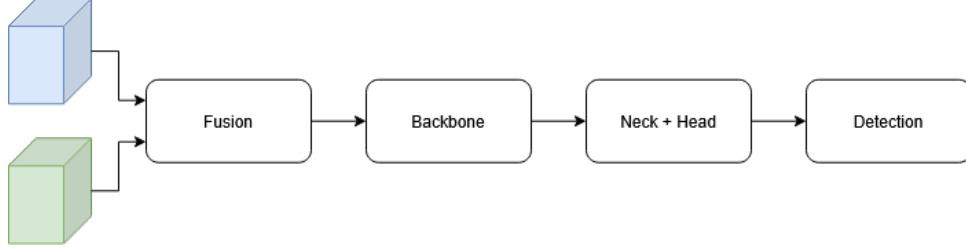
YOLOv5 [10] was selected owing to its widespread adoption, strong performance, and ease of customization. Our objective was to develop a generic, modular, multi-stream adaptation of YOLOv5 capable of processing an arbitrary number of input images with minimal modifications to the underlying codebase. To this end, the YOLO YAML configuration file was extended with additional parameters — *fusion_type*, *ni* (number of inputs), *fuse_steps*, and *fuse* — to specify fusion strategies and modules. The model is available in its GitHub repository.²

¹[Dataset](#)

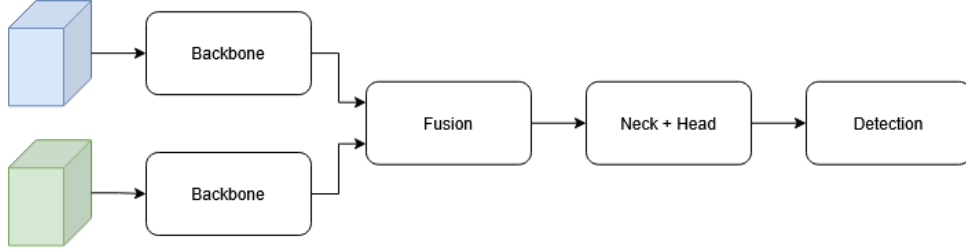
²[Model repository](#)



(a) YOLO method basics



(b) YOLO early fusion: fusion is the first operation of the network, before the backbone, which makes the overall network closer to a regular YOLO architecture



(c) YOLO late fusion: where fusion is done after the backbone, omitting the connections between the middle of the backbone and the fusion module for simplicity

Fig. 1: Simple diagrams of the different YOLO fusion methodologies

To accommodate multiple inputs, the input tensor for fusion networks was shaped as $[BatchSize, N * Depth, Width, Height]$, where N is the number of input images. This configuration enables the splitting and processing of individual inputs according to the requirements of different fusion strategies. Figure 1a illustrates the overall pipeline.

The following fusion methodologies were implemented: **Early Fusion**, in which a single fusion module is applied directly to the concatenated input images before the YOLO backbone (Figure 1b); and **Late Fusion**, where multiple backbones, one per input image, are employed, with fusion modules operating on outputs from specific backbone layers (*e.g.*, layers 4, 6, and 9) prior to passing the fused features to the neck and head (Figure 1c).

Transfer learning using pre-trained YOLOv5l³ weights was employed for both early and late fusion approaches to accelerate convergence and improve training efficiency.

Several fusion models were implemented: 1) *CatFuse*, a straightforward module that concatenates inputs along the depth dimension and applies a 1×1 convolution to unify the fused representation (Figure 2a); 2) *CBAMC*, an adaptation of the Convolutional Block Attention Module (CBAM), which enhances relevant features from each input using attention mechanisms before concatenation (Figure 2b); and 3) *TransEnc*, a transformer encoder-based fusion model employing patch embeddings and positional encodings to merge the incoming features effectively (Figure 2c).

³YOLOv5l is the "large" variant of the YOLOv5 object detection architecture, offering a good trade-off between accuracy and inference speed.

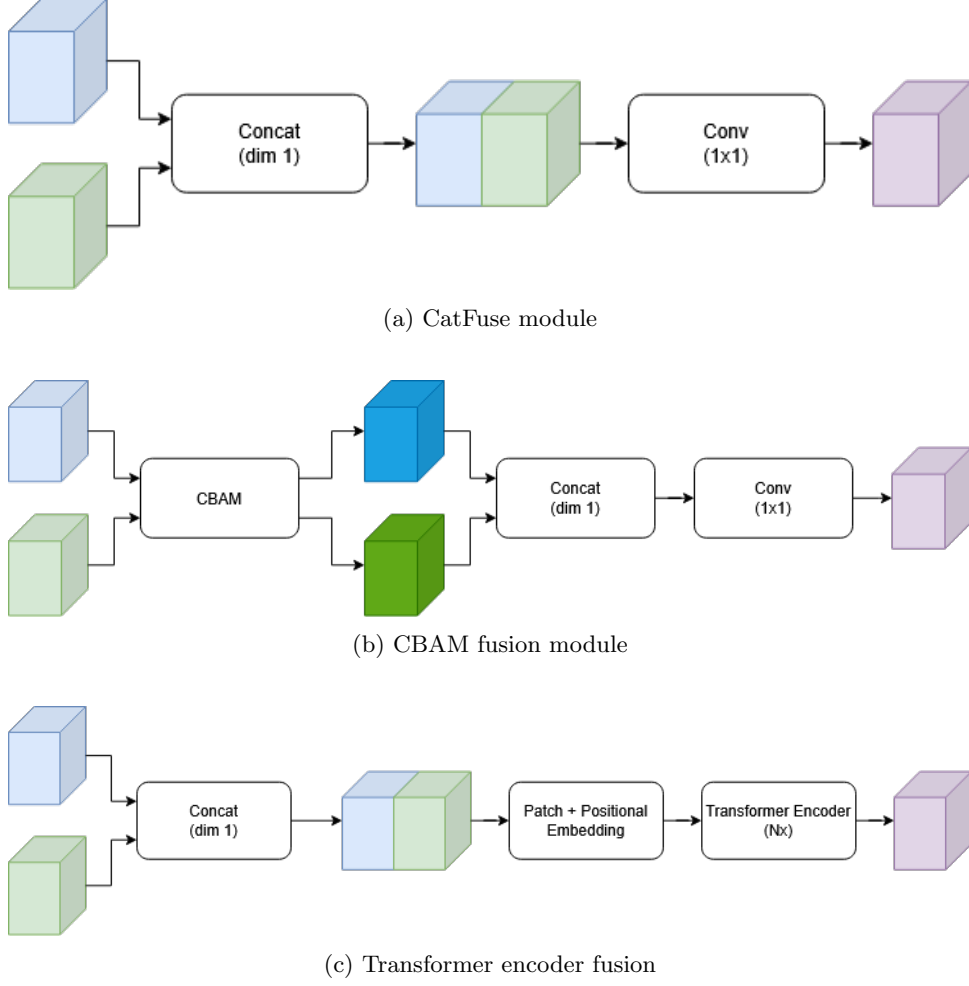


Fig. 2: Fusion module architectures: (a) CatFuse, (b) CBAMC, and (c) TransEnc.

2.3 Experimental Procedure and Analysis

Models were trained for 1000 epochs with early stopping triggered by a patience of 100 epochs, using a batch size of 4. Training and evaluation were conducted on a system equipped with an NVIDIA GeForce RTX 4060 12GB GPU. Baseline models were trained separately on single-image inputs corresponding to HMI, AIA 211 Å, and AIA 335 Å data.

Model performance was evaluated using mean Average Precision (mAP) at an Intersection over Union threshold of 0.5 (mAP@0.5) and the more stringent mAP averaged over thresholds from 0.5 to 0.95 (mAP@0.5:0.95). To rigorously compare model efficacy, we performed statistical analysis on per-instance F1 scores, which balance precision and recall. The Friedman test [14] was employed to detect overall significant differences among models, followed by Wilcoxon signed-rank post-hoc tests [15] for pairwise comparisons.

3 Results and Discussion

The baseline models were trained using YOLOv5l, initialized with pretrained COCO weights⁴. Each model was trained separately using a single image type (HMI Ic, AIA 211 Å, or AIA 335 Å) from the constructed multispectral dataset. The objective was to assess how different wavelengths of the electromagnetic spectrum individually influence sunspot detection performance.

The results are summarized in Table 1.

⁴COCO (Common Objects in Context Dataset) Dataset <https://cocodataset.org/>

Image Type	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Training Time (h)
HMI	0.89	0.88	0.92	0.50	1.325
AIA 211 Å	0.63	0.61	0.62	0.23	1.289
AIA 335 Å	0.69	0.54	0.58	0.24	1.532

Table 1: Baseline results

Module	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Training time (h)
Baseline (HMI)	0.89	0.88	0.92	0.50	1.325
Early CatFuse	0.90	0.90	0.94	0.52	2.156
Early CBAMC	0.90	0.89	0.94	0.53	2.693
Early TransEnc	0.84	0.81	0.85	0.41	3.667
Late CatFuse	0.87	0.76	0.79	0.39	5.031
Late CBAMC	0.86	0.79	0.80	0.40	4.815

Table 2: Results of all fusion methodologies developed and tested across both early and late methodologies

The HMI images produced the strongest results, achieving a mAP@0.5 of 0.92, which is consistent with their widespread use in sunspot detection research. In contrast, AIA 211 Å (mAP@0.5 of 0.62) and AIA 335 Å (mAP@0.5 of 0.58) yielded lower performance. Nevertheless, these results still demonstrate the model’s ability to learn relevant sunspot features from alternative wavelengths. Training times for these baseline models ranged between 1 and 1.5 hours, indicating that even the simplest single-image configurations can serve as effective starting points for further exploration. These findings underscore the significant influence of image type on detection performance, with HMI consistently proving to be the most effective for this task.

We then implemented and tested two primary fusion methodologies, *early fusion* and *late fusion*, with three distinct fusion modules: **CatFuse**, **CBAMC**, and **TransEnc**. The fusion results are presented in Table 2.

The fusion models yielded only modest improvements over the baselines, with the best cases achieving approximately a 4% increase in mAP@0.5 and 3% in mAP@0.5:0.95. Early fusion methods, particularly the **CatFuse** and **CBAMC** modules, demonstrated the most promising performance, both attaining a mAP@0.5 of 0.94. This is notable given the relative architectural simplicity of these modules. Conversely, the **TransEnc** module, which employs a more sophisticated attention mechanism, underperformed in its early fusion implementation, achieving a mAP@0.5 of only 0.85.

Late fusion approaches, despite their increased architectural complexity and more frequent fusion operations, yielded lower mAP@0.5 scores (0.79 for CatFuse and 0.80 for CBAMC) than both early fusion and the HMI baseline. Additionally, late fusion incurred substantially longer training times—approximately twice that of early fusion—further questioning its practical utility in this context. These results indicate that while multimodal fusion can potentially improve sunspot detection, the timing and method of fusion critically influence performance, with early fusion clearly outperforming late fusion strategies.

3.1 Statistical Analysis

To rigorously evaluate whether the observed performance differences between models are statistically significant, we performed the Friedman test on per-instance F1 scores. The test yielded an extremely low p -value (7.12×10^{-237}), strongly rejecting the null hypothesis that all models perform equivalently. This indicates that the differences in performance are not due to random chance.

Following this, the Wilcoxon post-hoc test was applied for pair-wise comparisons between all model pairs, with results summarized in Table 3. The matrix of p -values reveals several clear patterns:

- **Best-performing group:** The baseline HMI model, Early CatFuse, and Early CBAMC exhibit high p -values (close to or equal to 1.00) when compared among themselves, indicating no statistically

	Base HMI	Base AIA 211 Å	Base AIA 335 Å	Early CatFuse	Early CBAMC	Early TransEnc	Late CatFuse	Late CBAMC
Base HMI	1.00	$2.57e^{-74}$	$2.00e^{-78}$	0.67	0.67	$3.52e^{-13}$	$2.31e^{-12}$	$8.97e^{-11}$
Base AIA 211 Å	-	1.00	$1.72e^{-04}$	$2.80e^{-72}$	$6.26e^{-74}$	$5.25e^{-33}$	$5.55e^{-27}$	$7.48e^{-31}$
Base AIA 335 Å	-	-	1.00	$9.94e^{-79}$	$3.41e^{-82}$	$5.32e^{-44}$	$4.21e^{-39}$	$1.80e^{-42}$
Early CatFuse	-	-	-	1.00	1.00	$2.79e^{-29}$	$4.85e^{-21}$	$2.79e^{-15}$
Early CBAMC	-	-	-	-	1.00	$3.40e^{-30}$	$7.13e^{-22}$	$1.69e^{-15}$
Early TransEnc	-	-	-	-	-	1.00	1.00	1.00
Late CatFuse	-	-	-	-	-	-	1.00	0.69
Late CBAMC	-	-	-	-	-	-	-	1.00

Table 3: Wilcoxon post-hoc test results

significant differences within this group. However, comparisons of these models against others generally yield very low p -values, confirming their superior performance.

- **AIA baseline group:** Baseline AIA 211 Å and AIA 335 Å are statistically distinct from the best-performing group, as indicated by very low p -values when compared to them. However, their mutual comparison shows a relatively higher p -value, suggesting more similarity between these two models.
- **Lower-performing group:** Early TransEnc, Late CatFuse, and Late CBAMC form a group characterized by high p -values among themselves (ranging from 0.69 to 1.00), indicating no significant differences within this cluster but clear statistical separation from the best-performing models.

This statistical analysis corroborates the performance metrics, highlighting that simpler early fusion methods achieve significantly better and more consistent results than more complex late fusion or single-spectrum baselines.

Visual inspection of the detection results (Figures 3, 4, and 5) further corroborates the statistical analysis. The late fusion approaches notably struggled, exhibiting poorer detection accuracy, while the Early TransEnc model showed deficiencies, including missing detections and difficulty in accurately delineating larger sunspot structures. Overall, the effectiveness of advanced multispectral object detection techniques for sunspot detection varied considerably depending on the fusion methodology employed. In particular, transformer encoder-based fusion alone did not yield performance improvements in this application.

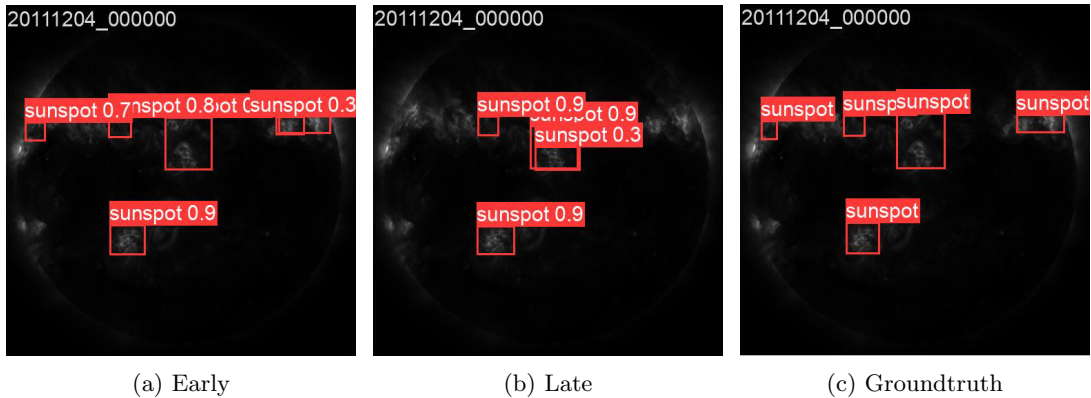


Fig. 3: Detection results of the CatFuse module in both early and late scenarios, alongside labels, where we can see the late model struggling to detect and the early model getting an almost perfect detection

Method	Description	Results
[8]	Faster R-CNN adaptation for feature fusion	≥ 0.9 precision/mAP@0.5 (dataset dependent)
[6]	Single wavelength YOLOv5 with Spectroheliograph H-Alpha Images	0.9 mAP@0.5
[19]	Combining different deep learning networks for single wavelength detection (HMI)	0.6582 Precision
[20]	Multi-wavelength Faster R-CNN using different AIA wavelengths (193 Å, 171 Å, 131 Å)	0.3 mAP@0.5

Table 4: Results comparison with other state-of-the-art models

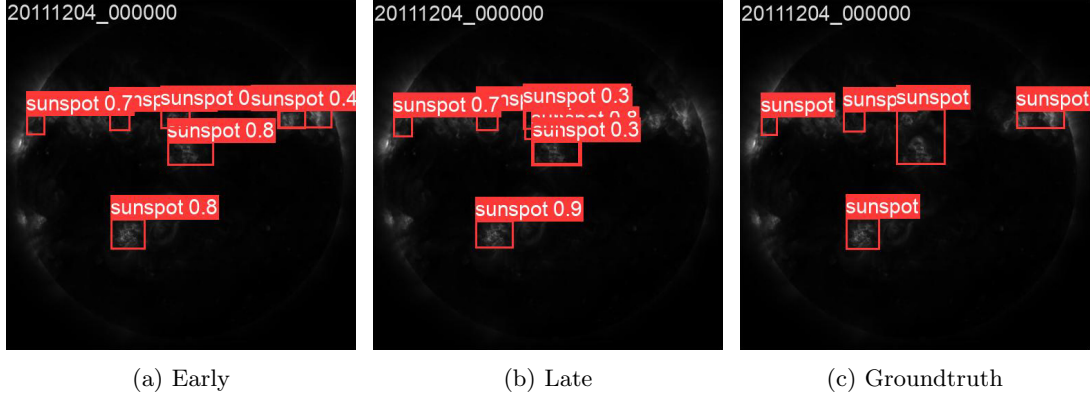


Fig. 4: Detection results of the CBAMC module in both early and late scenarios, alongside labels, where we can see the early model outperforming the late variant, while both struggle to match the results of CatFuse

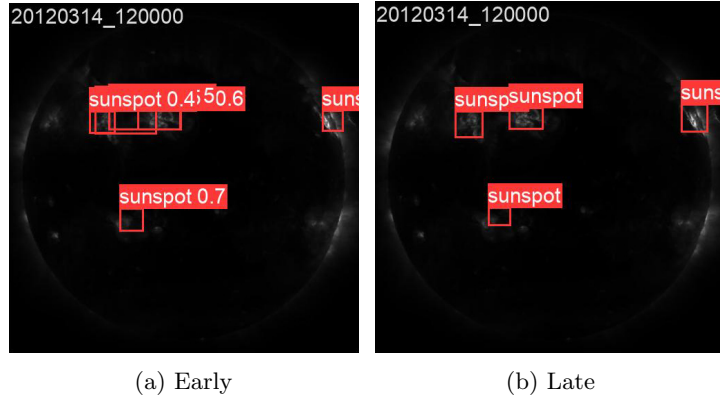


Fig. 5: Detection results of the early TransEnc approach, alongside labels, where we can see missing detections and difficulty in capturing the larger sunspot structure

While a direct and definitive comparison is complicated by differences in datasets and experimental setups, the results obtained in this study (best mAP@0.5 of 0.94) are generally consistent with the performance expected from contemporary deep learning methods applied to sunspot detection. Table 4 demonstrates that the developed models are competitive with existing approaches, despite the limitations posed by the lack of a standardized benchmark. These findings suggest that modern deep learning computer vision techniques can provide modest but meaningful improvements in sunspot detection by effectively leveraging multispectral imagery across different wavelengths.

4 Conclusion and Future Work

The primary aim of this work was to investigate how modern deep learning computer vision techniques can enhance sunspot detection by simultaneously leveraging images from multiple wavelengths and employing novel fusion methods to combine their features. The findings indicate that these techniques, particularly within the multispectral object detection (MSOD) framework, can yield a modest but positive impact on sunspot detection performance, with improvements of approximately 3% in the best cases. We successfully developed and adapted attention mechanisms, such as the Convolutional Block Attention Module (CBAM), for feature fusion and validated these approaches through statistical analysis. The proposed methodology performed at least as effectively as using a single HMI image in a standard object detection model and demonstrated competitiveness with other state-of-the-art sunspot detection methods.

Regarding the influence of solar image types, the evaluation across different electromagnetic wavelengths revealed that HMI images are generally the most effective for sunspot detection, consistent with existing literature. Although AIA 211 Å and AIA 335 Å images showed lower individual performance than HMI, their results indicate the network's capacity to learn relevant sunspot patterns from alternative wavelengths. The combination of multispectral images influenced model accuracy, with early fusion approaches using the CatFuse and CBAMC modules yielding the most notable benefits. In contrast, late fusion methods underperformed, achieving lower mAP@0.5 values compared to early fusion and even the HMI baseline.

When applying advanced MSOD techniques, fusion modules based on transformer encoders (TransEnc) in an early fusion setting did not improve performance relative to simpler fusion modules such as CatFuse and CBAMC. Additionally, the late TransEnc module was not feasible due to GPU memory constraints. These results suggest that transformer encoder-based fusion, without further task-specific refinement, may not inherently enhance model performance for sunspot detection.

Future work could focus on expanding the dataset with additional sources and exploring more specialized fusion architectures. Our experiments used a basic YOLOv5 backbone; therefore, improvements to the underlying network could further enhance performance, albeit requiring careful ablation to isolate the effects of fusion strategies. Due to memory constraints, the Late TransEnc module was excluded from the experiments as it exceeded the GPU's 12GB capacity as such, testing it with a bigger capacity and dataset could be advantageous. However, a bigger dataset would also be necessary.

This research lays a foundation for multispectral object detection in solar imaging and highlights promising avenues such as integrating image data with numerical or tabular satellite measurements to support multi-modal and physics-informed computer vision approaches [21]. The methodology could also be extended to multi-class detection and classification tasks, encompassing various solar phenomena like sunspots and solar flares concurrently, which would demand robust dataset curation and potentially automated annotation generation. Another promising direction is multispectral multiple object tracking, to track sunspot evolution over temporal sequences, despite current dataset limitations.

Funding Declaration. This work was financially supported by UI/BD/153350/2022 of the LIACC - Artificial Intelligence and Computer Science Laboratory - funded by Fundação para a Ciência e a Tecnologia, I.P./ MCTES through the national funds. It was also partially funded by the Portuguese Foundation for Science and Technology (FCT, Portugal), under the research grants UID/04434/2025.

References

- [1] Weber, M.A., Schunker, H., Jouve, L., Işık, E.: Understanding active region origins and emergence on the sun and other cool stars. *Space Science Reviews* **219**(8), 63 (2023)
- [2] Liewer, P.C., Neugebauer, M., Zurbuchen, T.: Characteristics of active-region sources of solar wind near solar maximum. *Solar Physics* **223**, 209–229 (2004)
- [3] Talib, M.A., Majzoub, S., Nasir, Q., Jamal, D.: A systematic literature review on hardware implementation of artificial intelligence algorithms. *The Journal of Supercomputing* **77**(2), 1897–1938 (2021)
- [4] Gupta, A.: Current research opportunities for image processing and computer vision. *Computer Science* **20**, 387–410 (2019)

- [5] Carvalho, S., Gomes, S., Barata, T., Lourenço, A., Peixinho, N.: Comparison of automatic methods to detect sunspots in the coimbra observatory spectroheliograms. *Astronomy and Computing* **32**, 100385 (2020)
- [6] Santos, J., Peixinho, N., Barata, T., Pereira, C., Coimbra, A.P., Crisóstomo, M.M., Mendes, M.: Sunspot detection using yolov5 in spectroheliograph h-alpha images. *Applied Sciences* **13**(10), 5833 (2023)
- [7] Abd, M.A., Majed, S.F., Zharkova, V.: Automated classification of sunspot groups with support vector machines. In: *Technological Developments in Networking, Education and Automation*, pp. 321–325 (2010). Springer
- [8] Almahasneh, M., Paiement, A., Xie, X., Aboudarham, J.: Msmt-cnn for solar active region detection with multi-spectral analysis. *SN Computer Science* **3**(3), 197 (2022)
- [9] Gong, L., Yang, Y., Feng, S., Dai, W., Liang, B., Xiong, J.: Solar active regions detection and tracking based on deep learning. *Solar Physics* **299**(8), 121 (2024)
- [10] Inc, U.: Ultralytics. <https://www.ultralytics.com/> Accessed 2025-06-30
- [11] Jocher, G., Stoken, A., Borovec, J., Changyu, L., Hogan, A., Diaconu, L., Ingham, F., Poznanski, J., Fang, J., Yu, L., et al.: ultralytics/yolov5: v3. 1-bug fixes and performance improvements. Zenodo (2020)
- [12] Gallagher, J.E., Oughton, E.J.: Surveying you only look once (yolo) multispectral object detection advancements, applications and challenges. *IEEE Access* (2025)
- [13] Baek, J.-H., Kim, S., Choi, S., Park, J., Kim, J., Jo, W., Kim, D.: Solar event detection using deep-learning-based object detection methods. *Solar Physics* **296**(11), 160 (2021)
- [14] Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research* **7**(Jan), 1–30 (2006)
- [15] Wilcoxon, F.: Individual comparisons by ranking methods. *Biometrics Bulletin* **1**(6), 80–83 (1945) <https://doi.org/10.2307/3001968>
- [16] Barnes, W.T., Bobra, M.G., Christe, S.D., Freij, N., Hayes, L.A., Ireland, J., Mumford, S., Perez-Suarez, D., Ryan, D.F., Shih, A.Y., et al.: The sunpy project: Open source development and status of the version 1.0 core package. *The Astrophysical Journal* **890**(1), 68 (2020)
- [17] Glogowski, K., Bobra, M.G., Choudhary, N., Amezcua, A.B., Mumford, S.J.: drms: A python package for accessing hmi and aia data. *Journal of Open Source Software* **4**(40), 1614 (2019)
- [18] Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. *International journal of computer vision* **88**, 303–338 (2010)
- [19] Palladino, L., Ntagiou, E., Klug, J., Palacios, J., Keil, R.: Sunspot groups detection and classification on sdo/hmi images using deep learning techniques. In: *2022 IEEE Aerospace Conference (AERO)*, pp. 1–10 (2022). IEEE
- [20] Kucuk, A., Aydin, B., Angryk, R.: Multi-wavelength solar event detection using faster r-cnn. In: *2017 IEEE International Conference on Big Data (Big Data)*, pp. 2552–2558 (2017). IEEE
- [21] Banerjee, C., Nguyen, K., Fookes, C., George, K.: Physics-informed computer vision: A review and perspectives. *ACM Comput. Surv.* **57**(1), 17–11738 (2024) <https://doi.org/10.1145/3689037>