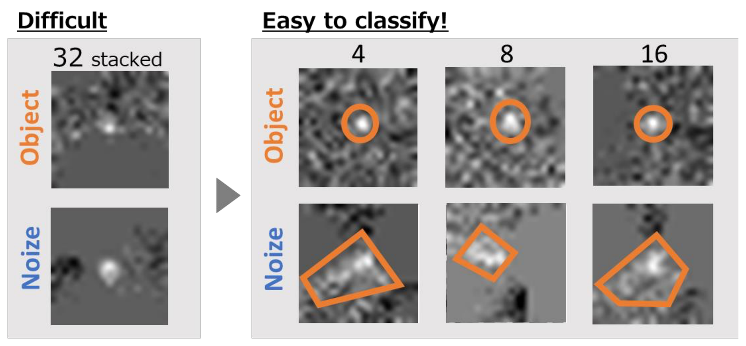
Multi-Input Convolutional Neural Networks with Attention for Faint Moving Object Detection

○ Masato SHIBUKAWA (SOKENDAI),
Fumi YOSHIDA (University of Occupational and Environmental Health・Chiba Institute of Technology University), Toshifumi YANAGISAWA (Star Signal Solutions Co., Ltd.), Takashi ITO (CFCA/NAOJ), Hirohisa KUROSAKI (JAXA), Neo YAMASHITA (KWANSEI GAKUIN UNIVERSITY), Makoto YOSHIKAWA(JAXA), Kohki KAMIYA(JAXA), Ji-an JIANG(University of Science and Technology of China),Wesley FRASER,JJ KAVELAARS(Herzberg Astronomy and Astrophysics Research Centre・ University of Victoria),Susan BENECCHI(Planetary Science Institute),Anne VERBISCER(Southwest Research Institute・University of Virginia),O Hosei(The University of Tokyo), Akira Hatakeyama(SOKENDAI), Naoya Ozaki(JAXA)

Background : We are searching for **faint Kuiper Belt objects** by stacking images acquired with HSC.
Aim : Developed a novel object detection model **using multiple stacked images as input**.
Method : The model combines a Convolutional Neural Network with a **Convolutional Block Attention Module**.
Result : Achieved an **AUC of 0.99** in classification performance and a **99% reduction** in human inspection workload.

1. Introduction

- Conducted Subaru Telescope observations to search for flyby objects and Kuiper Belt objects for the New Horizons Kuiper Extended Mission (KEM).
- Numerous faint distant object candidates were detected in the acquired image data.
- Human inspection used to separate real objects from noise.
- Single-frame classification showed limited reliability.



Developed a model for object detection based on multiple stacked images
(Manuscript submitted and under review in PASJ)

2. Data Processing

(i) Observation and processing

Table2.1 Observation Datasets Configuration

Parameter	Dataset 1	Dataset 2
DATEOBS (UT)	2020/06/20	2020/05/30
Red. id	03093	03072
Start time of the first exposure	10:55:39.175	10:31:55.756
Start time of the last exposure	14:55:24.839	14:50:04.288
RA2000 (°)	288.662	288.714
DEC2000 (°)	−20.049	−20.380
Field id	F2	F2
Filter	$r2$	$r2$
Exp time (s)	90	90
Number of images	118	128
Night condition	(Sky) high clouds, (Seeing) 0″7 – 1″1 (Temp) 6.9 – 7.6°C (Wind) 1.9 – 7.7 m s^{-1} (Humidity) 2.2%-24.4%	(Sky) clear, (Seeing) 0.″7-1″.2 (Temp) 3.6 – 5.9°C, (Wind) 2.5 – 15.4 m s^{-1} (Humidity) 17.5% – 28.5%

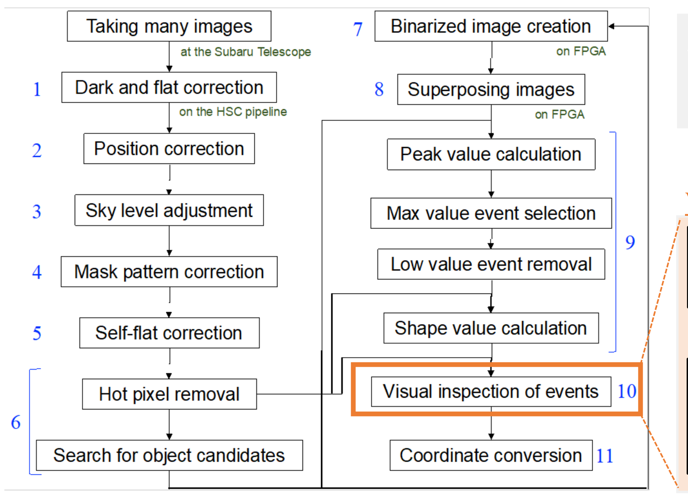


Fig2.1 Process of the JAXA system

Previous human tasks
We classified about 2,000 output images per observation.

We automate human tasks using AI.

Generate input data from an output of the JAXA system→(ii)

Classify the input data has object or noise.→(3. Machine Learning Model)

- Used two-night observation data obtained in 2020 at the Subaru Telescope with HSC[Yoshida+ (2024), Fraser+ (2024)].
- Applied image stacking with the JAXA Moving Object Detection System (see the poster P7 by Yamashita+)

(ii) Input data creation process

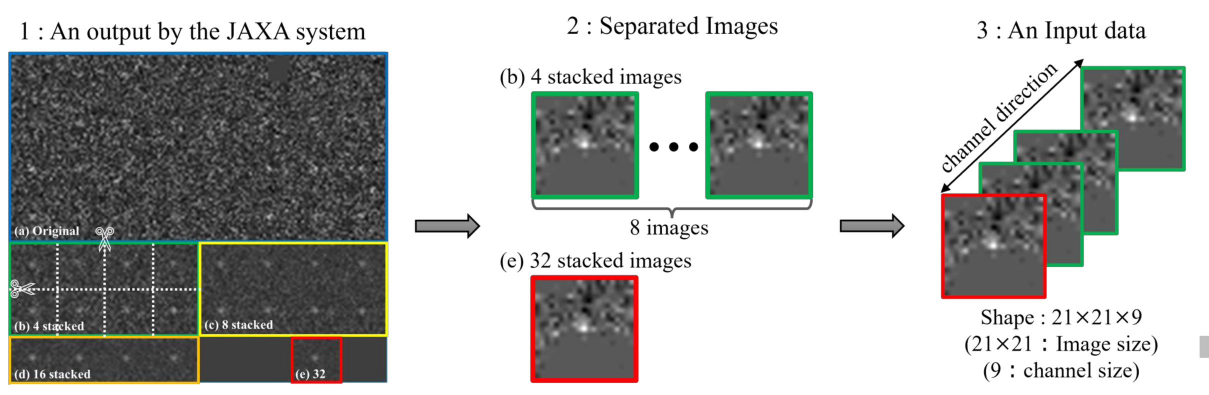


Fig. 2.2 Process of creating input images from the JAXA system outputs

Table2.2 Machine Learning Datasets Configuration

Feature	Dataset 1	Dataset 2
Red. id	03093	03072
Primary Use	Training	Evaluating
Total Stacked Images	1,966 (11,796)	3,015
Images with Objects	1,476 (8,856)	2,643
Images without Objects	490 (2,940)	372

Numbers in parentheses represent data after sixfold augmentation (e.g., 1,966 raw images became 11,796 after stacking).

Sixfold data augmentation by 90, 180, and 270 rotations, and vertical and horizontal flips.

- Created input data by stacking multiple superposing images.
- Applied rotation and flip augmentation to generate ~14,000 samples.

3. Machine Learning Model

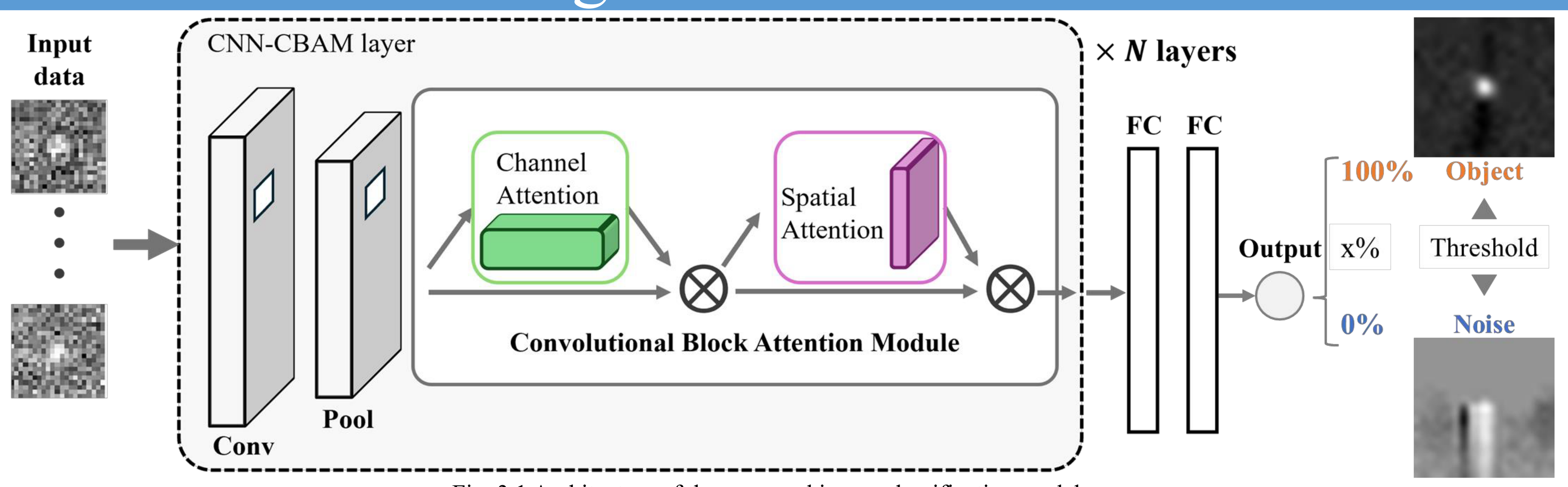
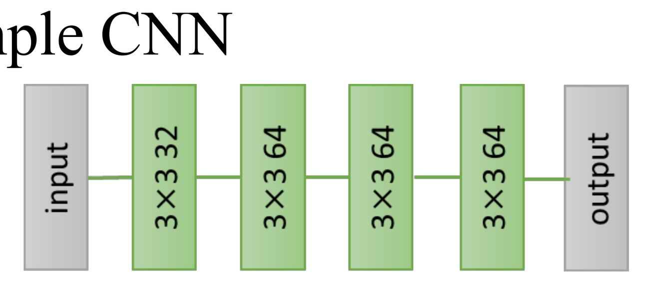


Fig. 3.1 Architecture of the proposed image classification model

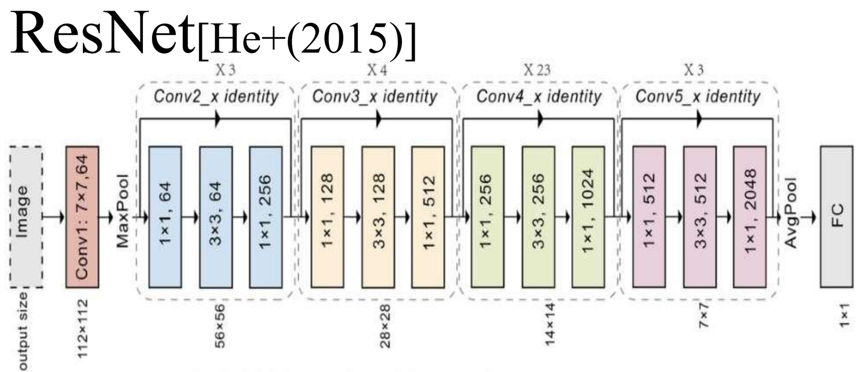
- Constructed a model combining a Convolutional Neural Network (CNN) and a Convolutional Block Attention Module (CBAM).
- Applied various backbone CNNs to explore architectures that best utilize CBAM performance in Simple CNN and ResNet.

Simple CNN



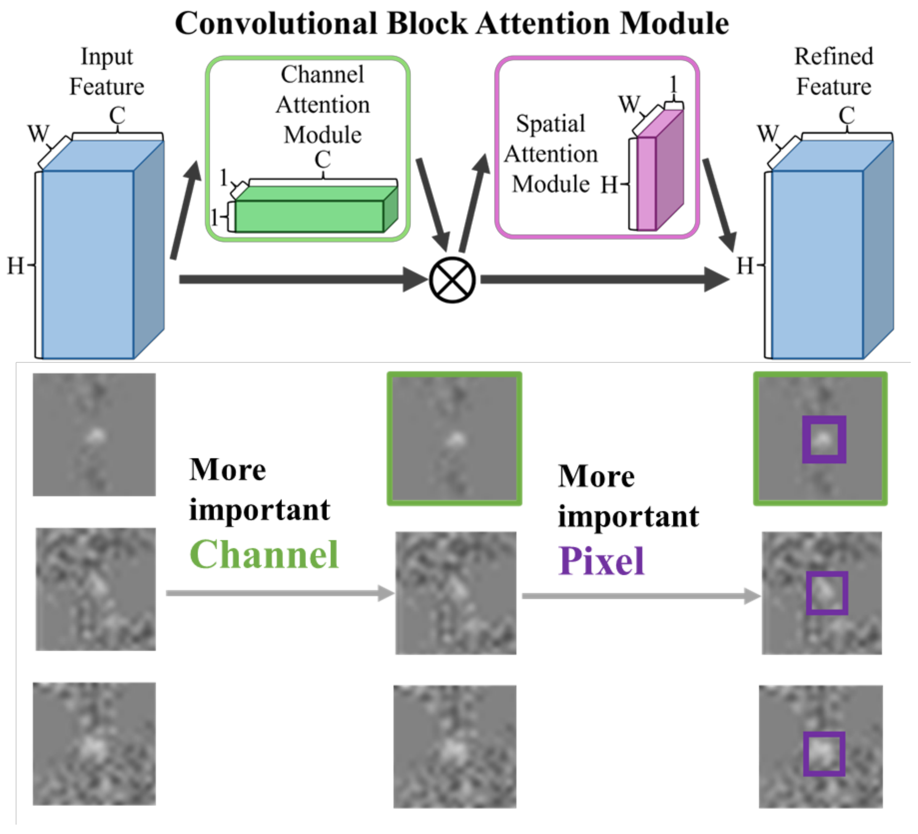
・Simple CNNs have 2 or 4 layers
・These has 6 different model with channels

ResNet[He+(2015)]



・We used ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152.
・ResNet is a deep CNN architecture that stabilizes training via skip connections, addressing the vanishing gradient problem

CBAM[Woo+(2018)]



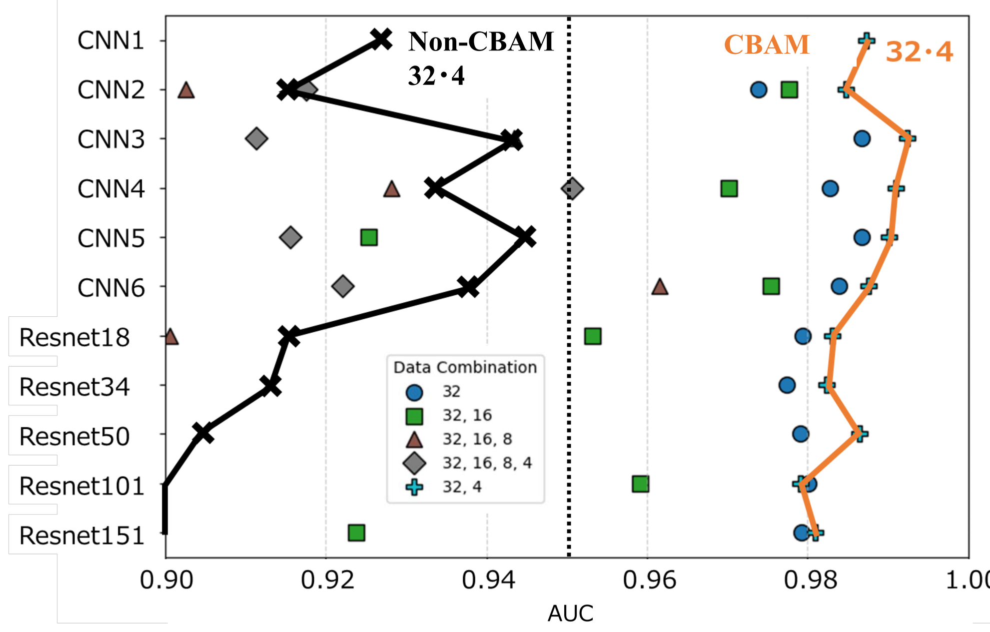
By learning both channel-wise and spatial attention, CBAM highlights important features within the feature maps.

4. Results

Table4.1 Model and Training configuration

Name	CNN1	CNN2	CNN3	CNN4	CNN5	CNN6	ResNet18	ResNet34	ResNet50	ResNet101	ResNet152
Channel set	32–64	64–128	32–32–64–64	64–64–128–128	32–64–128–256	64–128–256–512	-	-	-	-	-
Layer	2	2	4	4	4	4	18	34	50	101	152
Epoch	30										
Batch size	32										
Learning ratio	0.01										
Change ratio	0.1										
Change step	5, 10, 15, 20, 25										
Optimizer	SGD (Stochastic Gradient Descent)										
Loss function	BCE (Binary Cross Entropy)										
Momentum	0.9										
Weight decay	0.0001										

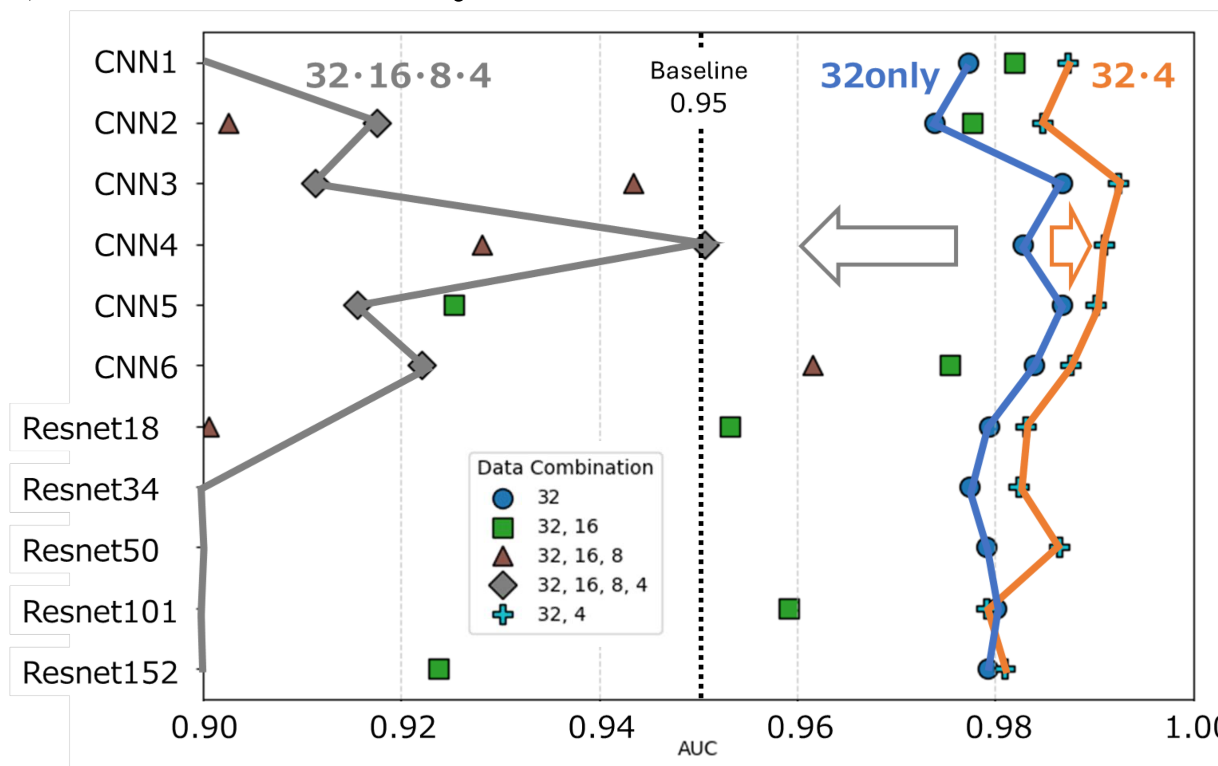
(i)CBAM model vs Non-CBAM model



【AUC (Area Under the Curve)】
A threshold-independent metric evaluating the classifier's overall ability to distinguish between classes.

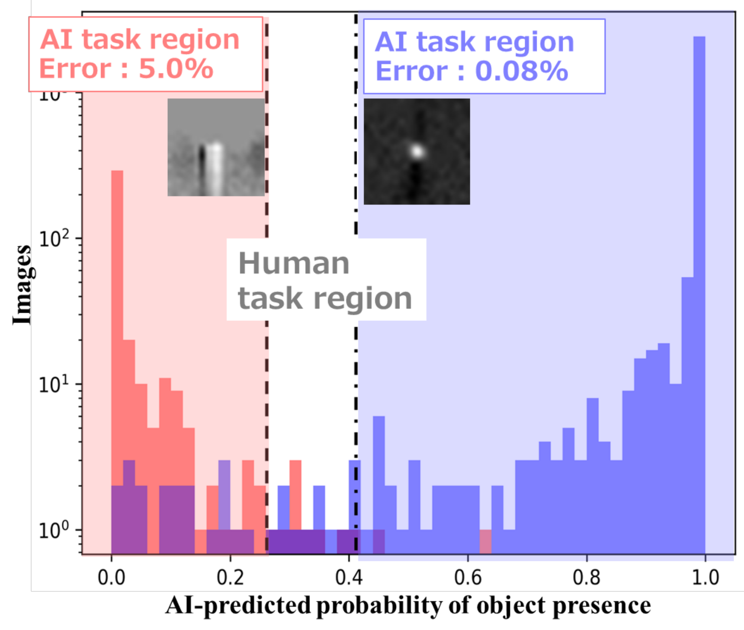
・**CBAM shows a clear improvement in AUC,**
・**AUC: 0.99 achieved**

(ii)Performance by Data Combination



・Combining 32・4 improved classification performance.
・All stacked images resulted in performance degradation.
・These results indicate that the combination of input images must be carefully optimized.

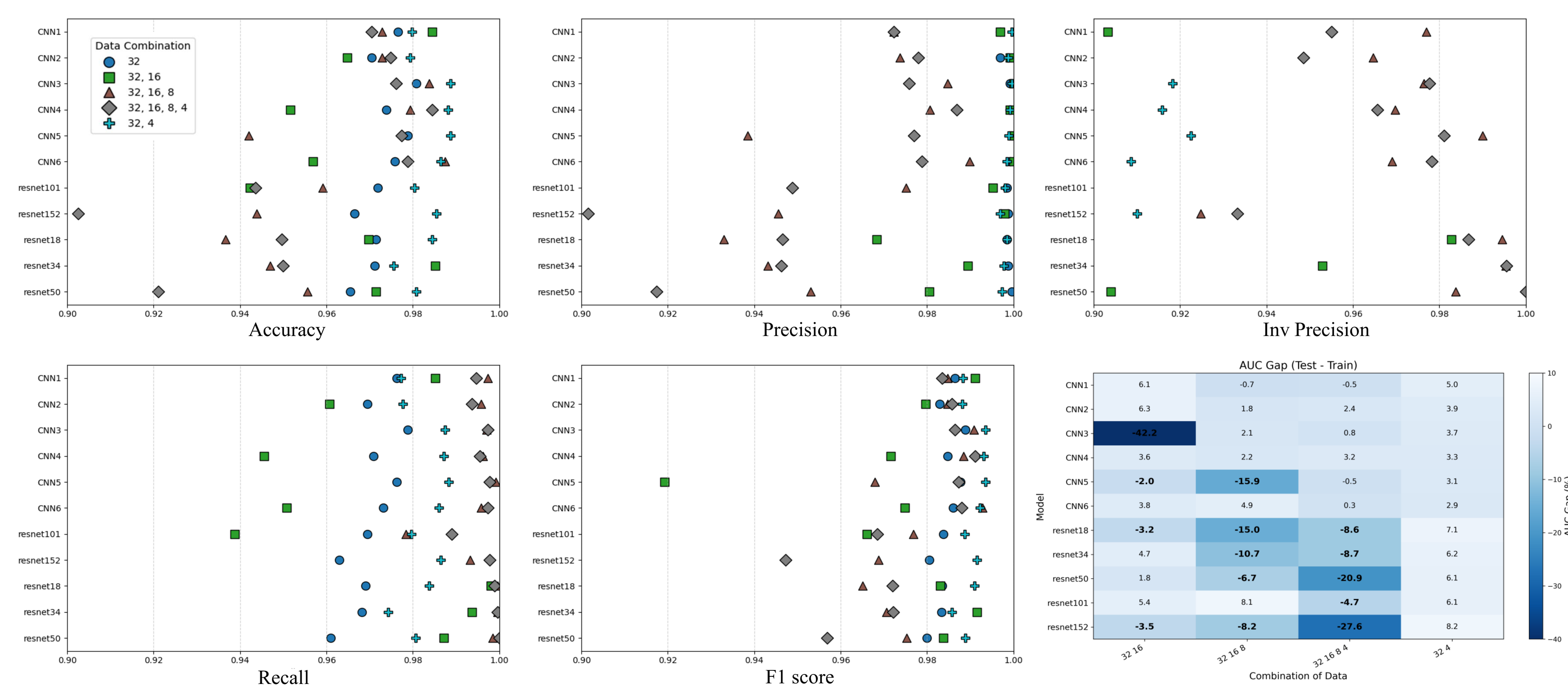
(iii)Result of Human task reduction ratio



Set a threshold on the model output to minimize human task rate while keeping AI error as low as possible.

・**Task reduction : 99.37%**
・**False detection :**
0.08% (with objects)
5% (without objects)

＊Results of Other Standard Metrics



This colormap represents the investigation of overfitting and darker colors indicate stronger activation.

5. Conclusion

- Classification Performance
  - Achieved an **AUC of 0.99** in classification performance.
  - Confirmed **improved classification accuracy with multi-image input**.
  - Observed performance **enhancement through CBAM integration**.
  - Achieved a **99% reduction** in human inspection workload.
- Future Work
  - Integration into the operational pipeline for new object discovery.
  - Improving model generalization to handle diverse observation images.