M²SODAI: Multi-Modal Maritime Object Detection Dataset With RGB and Hyperspectral Image Sensors

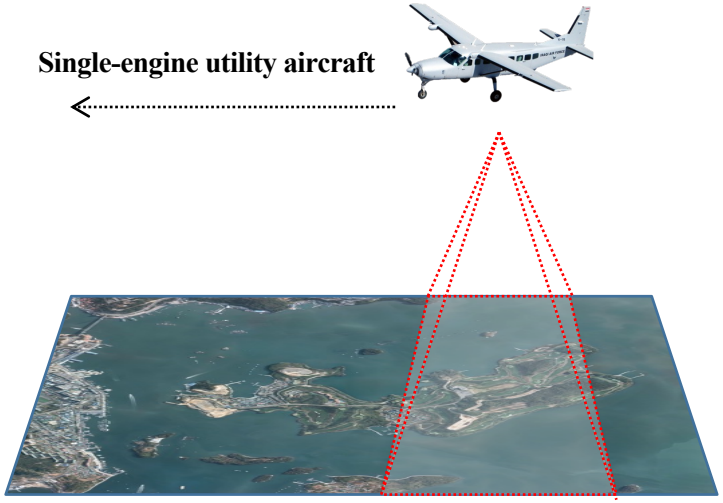
Jonggyu Jang, Sangwoo Oh, Youjin Kim,
Dongmin Seo, Youngchol Choi, Hyun Jong Yang*

NeurIPS 2023

Target Scenario: Aerial Object Detection in Maritime Space

1. Ship
2. Floating matter (buoys, rescue tubes, small lifeboats, surfboards, and humans)

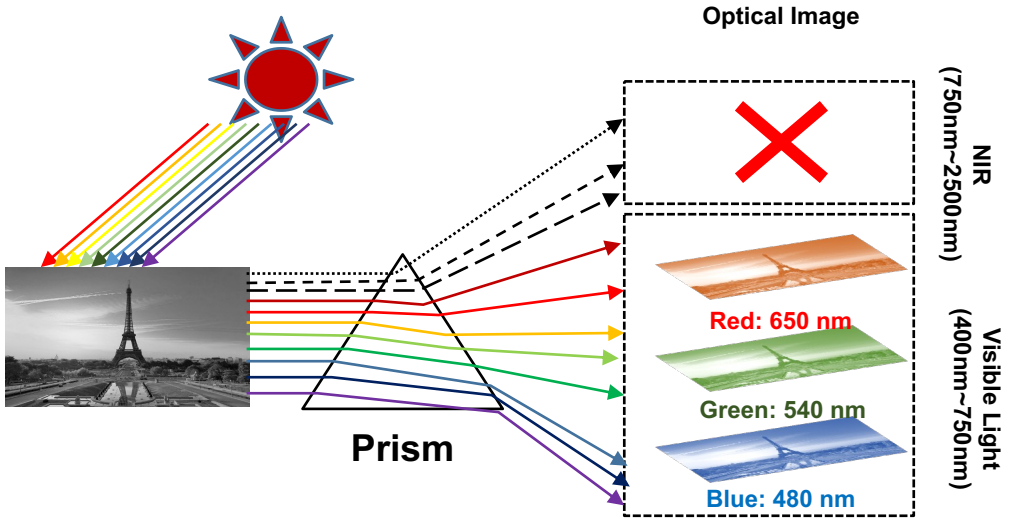
Aerial Object Detection Scenario

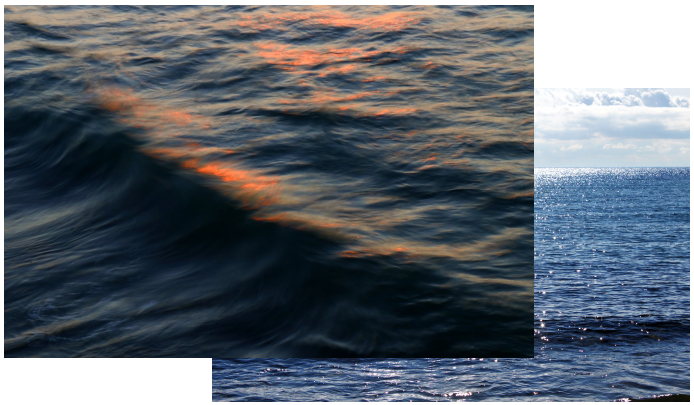


Object Detection Example

Motivation: RGB Sensor

Spatial resolution: High (0.1 m)
Spectral Information: Small (3 Channel)



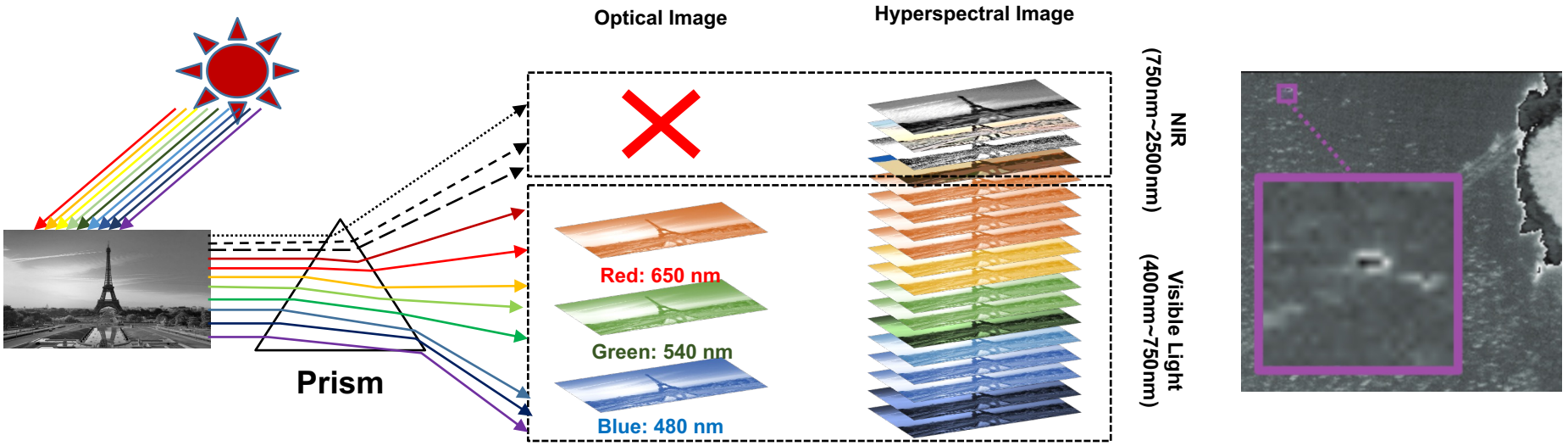


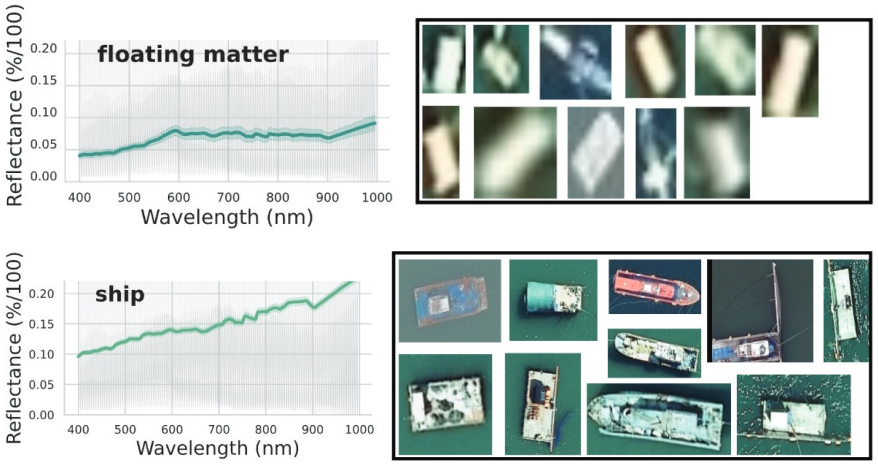
It is hard to recognize small objects if wave or sea glitters are on sea surface.

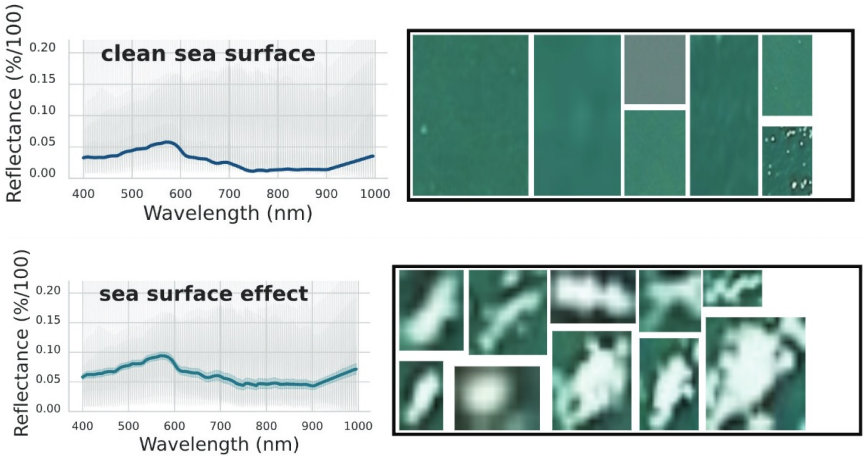
Challenges: Wave & Sea Glitter

Motivation: Hyperspectral (HSI) Sensor

Spatial resolution: Low (0.7 m)
Spectral Information: Large (<100 Channel)







Target Objects

Backgrounds

Objective: Multi-modal dataset

1. RGB sensor data
2. HSI sensor data

High resolution RGB sensor data
High spatial resolution
Low spectral resolution

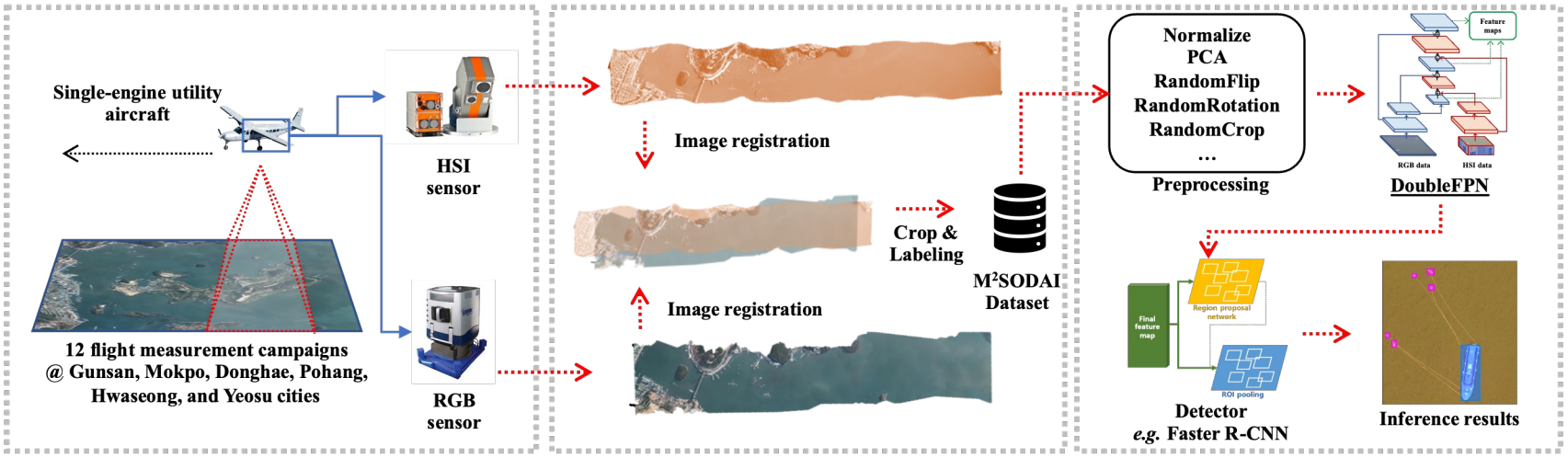
+

Low resolution HSI data
Low spatial resolution
High spectral resolution

❑ We use an additional sensor data (high-resolution RGB sensor)

Our Approach

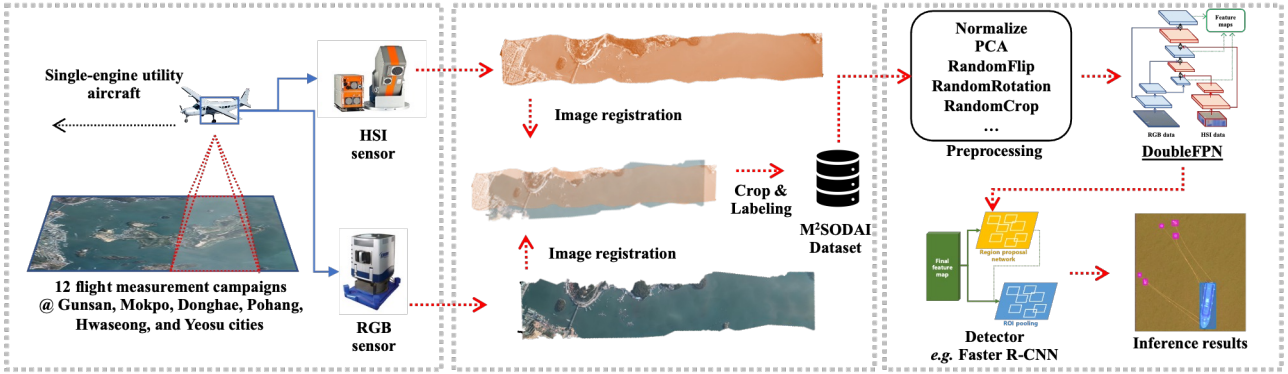
1) Data Collection
2) Image Registration
3) Experiments



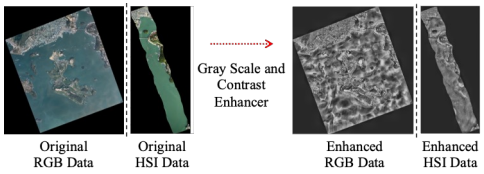
(a) Data collection (b) Image registration (b) Experiments

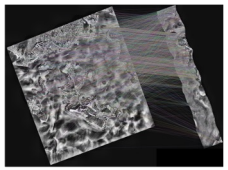
Our Approach

1) Data Collection
2) Image Registration
3) Experiments



2-1. The RGB and HSI sensor data are transformed by a gray scaler and a contrast enhancer.





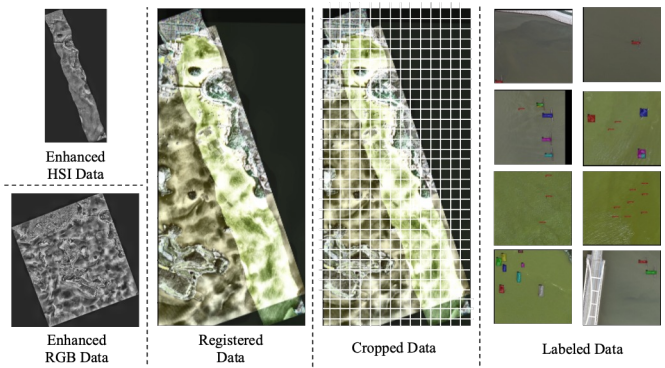
ORB Feature Extractor and Brute-Force Matcher

$$H = \begin{bmatrix} 1.02 \cdot 10^{0} & -2.87 \cdot 10^{-2} & -4.47 \cdot 10^{3} \\ 1.65 \cdot 10^{-2} & 1.00 \cdot 10^{0} & 2.74 \cdot 10^{2} \\ 6.47 \cdot 10^{-7} & -4.51 \cdot 10^{-7} & 1.00 \cdot 10^{0} \end{bmatrix}$$

Homography Matirx

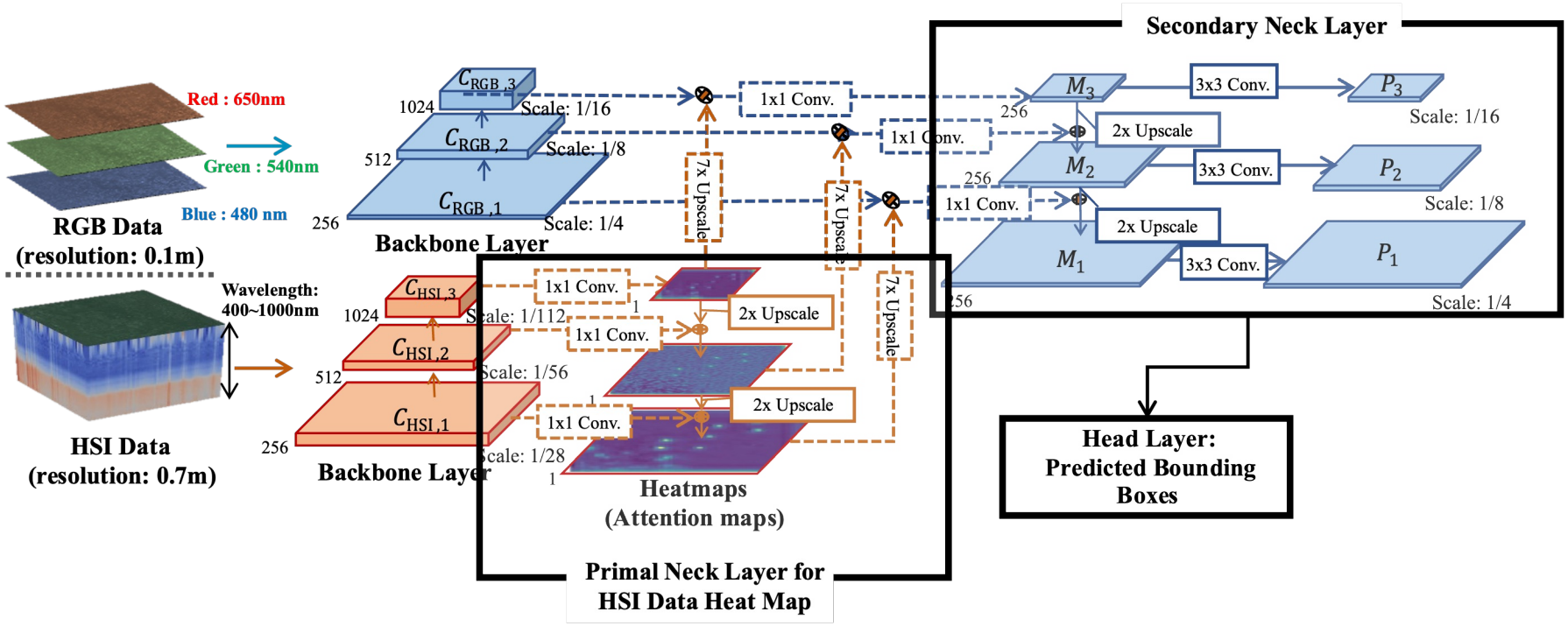
2.2. The matched feature of transformed sensor data and the corresponding homography matrix are obtained.

2.3. The sensor data are registered, cropped, and labeled.



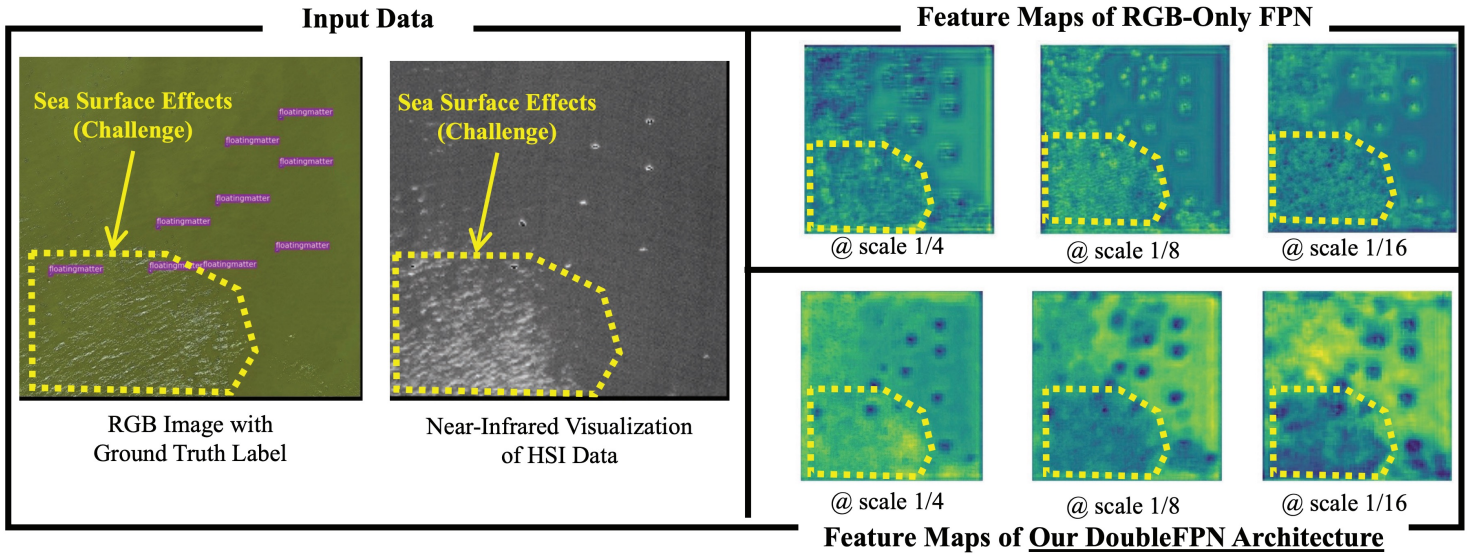
Our Approach

1) Data Collection
2) Image Registration
3) Experiments



Experimental Results

Information fusion gives more clear feature maps
Also, the proposed method has the highest performance compared to the baseline methods



neck layer	multi-modal	RGB data	HSI data	mAP	$AP_{@.5}$	$AP_{@.75}$	Ship	Float. Mat.	AP_s	AP_m	AP_l
DoubleFPN(ours)	✓	✓	✓	**44.4**	**84.8**	39.3	55.7	**33.1**	**35.2**	41.7	**61.4**
FPN (RGB) [26]	×	✓	×	38.8	77.0	33.3	52.4	25.2	18.3	**44.8**	55.6
FPN (HSI) [26]	×	×	✓	7.8	23.2	2.9	15.8	0.0	-	-	-
UA-CMDet [43]	✓	✓	✓	42.9	84.0	**40.0**	**55.9**	29.8	20.8	43.0	60.8
DetFusion [44]	✓	✓	✓	42.0	84.3	35.4	53.5	30.5	24.2	41.9	61.1
Early fusion	✓	✓	✓	42.9	83.0	37.6	54.2	31.5	18.9	44.1	59.7

Thank you 😀