This is my second review of this manuscript. I abandoned my first review after reading the first 200+ lines because I found the presentation to be confusing, incomplete, and in places incorrect, making it difficult to follow the authors' reasoning. I provide examples of each of these in my review. The authors have made a serious effort to address my concerns but, unfortunately, the revision falls short. There are two classes of problems: 1) questions about the procedures and 2) the clarity of the writing (organizational/editorial issues). I deal with each of these below:

Questions about Procedures

1) The abstract argues that no publicly "*available, longterm ocean front dataset currently exists, and the existing detection methods often rely on time-consuming manual labelling or traditional algorithms with limited accuracy in complex frontal regions."* This is true, but there have been a significant number of papers describing the global frontal field but, again, none that are publicly available. The important point is that the vast majority of previous work has been with fronts detected in L2 or L3 SST fields, not L4 fields but the dataset produced as part of this work is from L4 model reanalyses and, yet this is not mentioned in the abstract. I tend not to use L4 fields because of lost resolution and because of fronts resulting from artifacts in the field. This is an issue that is really not addressed in this manuscript.

2) It was not always clear to me what data you were using—critical to really understand what you did. In Section 2 you say that you will use the daily average sea-surface temperature (SST) output from GLORYS12v1. That's clear although it would be good to provide a DOI for the specific data you used as well as a date you downloaded it. However, in Section 3.2 you say "*To expand the dataset, satellite remote sensing images…".* This sentence sounds like you are using raw satellite SST fields, not the

GLORYS12v1 fields. If that is the case, what is the provenance of these fields? If not, this sentence should be rewritten. When you mention the preprocessing step (lines 174-175), it sounds like you are using raw satellite SST data, not the L4 GLORYS12v1 product. And in Section *4.5.1 Data source errors* you say "*The accuracy of ocean front detection is influenced by data quality. If the data include noise or missing values or the data source is inconsistent, the model will learn inaccurate features and introduce errors. The reliance on satellite imagery as the primary data source for the proposed method may limit its applicability in certain situations. In such cases, complementary data sources, such as in situ measurements or marine numerical model results, may enhance the accuracy and robustness of the model."* I'm confused here. Didn't you use a L4 model derived product not original satellite imagery? If so, what sort of noise are you referring to here and what gives rise to missing data. This section is important but it's really not clear where you're going with it. Are you suggesting the problems that might arise if one were to apply your algorithm to L3 SST data or are you discussing something else? This brings me back to the lack of clarity in places with regard to the data you are using. I think that the manuscript would benefit if you were more explicit.

3) Lines 174-175: "*2) Pretreatment: Preprocessing steps for the SST image data were conducted to enhance the quality and remove any noise or artefacts."* What do you mean by noise or artifacts here? You're using L4 fields right? What kind of noise did you detect in them and how did you remove this noise or these artifacts?

4) In your discussion of training dataset size (lines 242-258 and Table 3), it

sounds like you are varying the test dataset. Varying the train/test split is fine for studying data-efficiency, but the test set should not be used to select the split (or any other setting). Please clarify whether a separate validation set (or nested CV) was used for selecting hyperparameters/ split ratio, and whether the reported test set was held fixed and used only once for final evaluation. Also clarify what is meant by 'adding test data' improving generalization.

5) Lines 296-298: "*Compared with traditional methods, the ocean fronts extracted by the deep learning model have a higher degree of fit, and in addition, the deep learning model can better reflect the characteristics of ocean fronts."* I'm confused here. The gradient of the input data is shown on the left in the Fig. 8 and the gradient of the deep learning image is shown on the right. Isn't the gradient from the input image the actual correct value? The deep learning image is simulating a gradient, but it seems that you want that gradient to be close to the one observed from the real data but not necessarily equal to it. You need to clarify why the deep learning values are better if that's what you mean.

6) Line 307**:** You say that the width is defined as "distance from the centerline to the boundary," but it is not clear how the centerline could be defined and the distance to the boundary computed. Please add one or two sentences clarifying how the "centerline" is defined/computed from the front mask and how the distance to the boundary is measured (e.g., skeleton/medial axis + Euclidean distance transform, and whether the reported width is an average over centerline points). This would make the width metric fully reproducible.

Clarity of the Writing.

In my first review I suggested that the authors use a chatbot to help them with the English, specifically to address editorial/rhetorical problems. The authors responded that they used an editing service for this revision but this editing service does not appear to have addressed structure/clarity issues, which I think still plague the manuscript. Some examples:

1) Section 3.1 Gradient Calculation Method. First, it is not clear to me why the authors refer to this as the gradient calculation 'method', why not just the gradient calculation. Second, the paragraph contains extraneous material, which doesn't add much if anything to the story. Specifically, this section can be reduced to a few precise sentences defining the centered-difference derivatives and the gradient magnitude. For example,

"We estimate the SST gradient magnitude G from the gridded SST field T(i,j) using centered finite differences:

$$D_x(i,j)=\frac{T(i,j+1)-T(i,j-1)}{2\Delta x},\qquad D_y(i,j)=\frac{T(i+1,j)-T(i-1,j)}{2\Delta y},$$
$$G(i,j)=\sqrt{D_x(i,j)^2 + D_y(i,j)^2}.$$

Here Δx and Δy are the grid spacings (in km), and (i,j) denotes the grid indices."

Note that this is much shorter but, more importantly, it is to the point and adds clarity to the description with the specification that the spacing is in kms, assuming that it is. Also, note that I have fixed the incorrect subscript in Eq. 3.

2) Section 3.2 Data Labels.

This section is currently too vague to be reproducible. It introduces Labelme before explaining what it is, and it is unclear what constitutes one "annotation dataset." Please define the unit of annotation (one image/field? one tile?), specify whether multiple fronts can be annotated

per image, and clarify the source of the “satellite remote sensing images” used for dataset expansion (are these the same fields described in Section 2, or an additional data source?). Finally, state the output label format and what each JSON file corresponds to. Here’s a suggested revision of the section:

3.2 Data labels

Ocean fronts were manually delineated using *Labelme* (an open-source image-annotation tool) by drawing polygon boundaries around visually continuous front features in each SST-gradient map. Each annotated map was saved as a Labelme JSON file containing one or more polygons (multiple fronts may occur in a single map). We assembled a labeled dataset of 5,000 annotated SST-gradient samples (each sample = one SST-gradient map plus its corresponding JSON annotation file). To increase diversity, the samples were selected across different regions and seasons.

Of course, this is my understanding of what you have done so it may not be quite correct. If additional satellite remote-sensing products were used for this augmentation, the specific data sources and preprocessing steps should either be specified here or in another section, which you reference. Also, there are at least two versions of Labelme software so you should reference the one you use; e.g., if you use the Python tool, you might reference Wada, K. *labelme: Image Polygonal Annotation with Python* (computer software). Zenodo. https://doi.org/10.5281/zenodo.5711226

```
@software{wada_labelme_2021,
  author  = {Wada, Kentaro},
```

```
  title   = {labelme: Image Polygonal Annotation with
Python},
  year    = {2021},
  publisher = {Zenodo},
  doi     = {10.5281/zenodo.5711226},
}
```

But you may be using the MIT CSAIL LabelMe project/tool?

These are just examples of significant rhetorical modifications that could be made to improve the readability of the document. I used ChatGPT to help me structure these and, if the Journal allows you to use an LLM to improve your manuscript, again, I suggest that you do so.

Here are other editorial issues I noticed while reading the document:

3) Line 55: "with an mean Dice Similarity Coefficient (mDSC)" —> "with [a] mean Dice Similarity Coefficient (mDSC)" and a reference to 'Dice Similarity Coefficient' would be helpful.

4) Lines 126-128. I found the use of colour to be confusing here: "*However, detecting and characterizing these fronts is challenging due to their complex and dynamic nature. In particular, the visual similarities among different fronts can make it difficult to distinguish them based on colour and shape alone."* Please use consistent terminology—you use grayscale elsewhere in the manuscript—distinguishing the *data* (SST field/gradient field/raster) from *visualizations* (color-mapped images). If inputs are single-channel, avoid 'colour' and instead refer to 'intensity/value/ gradient magnitude' and spatial pattern/texture. For example, you might want something more like: "In particular, the visual similarity of different

fronts can make them difficult to distinguish based on gradient magnitude patterns and shape alone." if single channel gradients or "In particular, the visual similarity of different fronts can make them difficult to distinguish based on colormap intensity patterns and shape alone (noting that color is a visualization choice rather than a physical variable)." if colormap. And, if you truly use RGB composites (unlikely from what you've shown) then you need to state what the channels are.

5) Line 194: "*First, the accuracy of the border was represented..."* What do you mean by the border, is it the border of the image or the front you have found? It's actually pretty clear that it is not the edge of the image but up until now you have used front so, if that is what you mean, why not use it?

6) Line 220: You mention 'COCO-style instance labels'. I'm not familiar with these. You need at a minimum a parenthetical description accompanied with a reference.

7) Lines 229-230: "*Notably, to ensure the accuracy of the evaluation, the test set data do not need to participate in the training process."* What do you mean by the test step not needing to participate? Either they do or they don't. If the latter, which I'm assuming is the case, just say so: "…*the test set data do not participate in the training process."*

8) Lines 265-280: The first paragraph here is just a verbalization of what's in Table 4. You should only verbalize the parts of the table, which you emphasize in subsequent discussion. In fact, the 2nd paragraph discusses the table in more detail. These two paragraphs can be combined and shortened.

9) Line 324: "*Spring and summer have relatively fewer ocean fronts, with the...*" Fig. 9 doesn't show numbers but rather gradients. Yes, there are places where there are not gradients in spring and summer so there are less but, still, I would have an image with color being counts if the number is really important.

10) Lines 326-327 reference locations in Fig. 9. It would help for readers not very familiar with the region to label the locations mentioned in the figure.

11) Lines 354-356: I wouldn't use 'hydrological' here. In most scientific writing, hydrological points readers toward hydrology on land (precipitation, runoff, rivers, groundwater, watershed processes) and "the water cycle," not the ocean's water-mass properties. If you mean T/S/density and water masses, you are talking about hydrographic conditions. If you mean currents/front dynamics specifically, then use oceanographic or physical oceanographic conditions**.** I would make it **;** e.g., *"...Rapid changes in temperature and salinity gradients may affect..."* if that is appropriate.