Referee Comments on ContrailStruct30

This paper provides what I think is the world's best LEO-satellite dataset of per-pixel contrail labels. This is a valuable contribution that should be published after my comments below are addressed. I have one comment about the scientific substance of the paper and a few about the presentation.

Biases of the iterative label-generation scheme
This is my only comment about the substance of the paper. My understanding is that labels were iteratively created by a computer vision model, and edited by humans, with those labels then used to train the next generation of the computer vision model. I'm mostly worried about how this might produce biases in the result. For example, if the computer vision model erroneously decides that a class of images doesn't contain contrails, might it never show those images to the humans, so they never get labeled?

One obvious way to address this would be to run the computer vision model in a high-recall/low-precision mode, did the authors do this? If not, can they point to literature where others have done this kind of label generation?

The possible presence of these biases makes me question whether Section 5 really has any meaning. Basically the authors find that the model (MaxVit+Unet) that they used to generate the labels is also the best at classifying those labels. I suspect that if the authors had used e.g. SegFormer-B0 to assist the human labeling, then that model would have performed the best in Section 5. Similarly in section 5.5, the authors find that the color scheme used to do the labeling is the best one to pass to the computer vision models. This is not really surprising. Probably the contrails with high contrast in this color scheme are the ones that got labeled.

I think significant caveats should be added to Section 5 for this reason, or even consider removing it entirely.

I'd also like more details on how the iterative improvement worked. Were the models and humans iterating on the same patches or different ones. How did the size of the dataset being labeled change between iterations?

Assumed knowledge about computer vision
The authors are experts in computer vision but they are writing in an Earth Science journal which may not have the same familiarity with computer vision. Please explain a little more of the basics, and use less computer vision jargon. For example

- -include references for what a MaxViT–U-Net is
- -instead of "foreground/background", how about "positive/negative" or even "contrail/non-contrail"
- -what does "eight-connected" mean

- -in the sentence “The encoder is **timm's maxvit_base_tf_512.in21k_ft_in1k**, initialized with weights pretrained on **ImageNet 21K** and fine-tuned on **ImageNet-1K**” I don’t think the average reader knows what the bolded things are, can you please provide references and/or links
- -define the metrics in Fig 7/Table 6

Unclear/inconsistent language
The authors could make the text more readable by being more precise in their language.

For example, the authors say: “regions interpreted as contrails are foreground, and all other regions are background” and later “The two experts jointly selected difficult negatives from regions prone to human misinterpretation or 211 repeated model false detections”. What does ‘region’ mean in these two sentences. I think in the first sentence it means ‘pixel’, can you just say ‘pixel’. What does it mean in the second sentence?

What is a “scene”: I think a sentence saying e.g. “SDGSAT-1 takes images in X pixel by X pixel scenes, we dividing these into 256 x 256 pixel patches” would help

Highlight pixel-level labels
The big difference between this paper and Yu 2024 is that it has per-pixel annotations. This is pointed out in the paper but honestly I was almost at the end before I realized it. I have some ideas on how to make this more obvious to the reader.

Mention it more. In sentences like “The dataset's value for such studies is therefore its detailed morphology, which can be linked to additional evidence, rather than an age assignment based on shape alone.”, (and others that mention morphology) make it clear that this detailed morphology is available because of the per-pixel level labels

Be clearer in the ‘Annotations’ column in Table 1. I think every row should say either pixel masks, polygons, or image-level labels. Right now you point out if a row produces polygons or image-level labels, but don’t explicitly point out that all the others are per-pixel labels.

Minor comments

- The introduction (especially the references to Meijer, Ng and Santos Gabriel) should probably include a sentence or two discussing the differences and relative merits of contrail detection on LEO vs GEO satellites
- Should probably cite https://doi.org/10.5194/egusphere-2026-4608 which also does some machine-assisted contrail labeling
- I didn’t know what ADSB.1o1 was when I first read the text, I had to Google it. Maybe provide a citation or even just a hyperlink
- Fig3c legend: records -> patches

- Fig 6: I don’t see much difference between b/c and e/f. Is there supposed to be a contrail mask in the blue box?