Response to Review
Alexander Ukhov

I enjoyed reading this manuscript. The construction of a large SO2 plume database using a U-Net segmentation approach has value for the Community.

A minor concern is the reliance on a relatively small set of manually drawn plume masks (1000 plumes) for training. Manual plume labeling is very subjective.

One possible workaround is to generate plume masks using a Lagrangian dispersion model (e.g., FLEXPART-WRF driven by WRF winds).

We used FLEXPART-WRF in our recent work (Ukhov et al., 2025, JGR Atmospheres, https://doi.org/10.1029/2025JD043334) for SO2 point sources and emphasized that dense source clusters

and incomplete background removal can bias plume-based top-down, which is relevant to your discussion of regions where individual plumes cannot be isolated.

Question: how does your post-processing handle overlapping or merged plumes from closely spaced sources? For example, how do you avoid double-counting in the emission-rate estimate?

We thank the reviewer for their comments on our paper.

We agree that plume labelling is subjective and does result in the model only being as accurate as the person labelling the training data. However, we decided to use real data instead of modelled data to try and ensure we capture a full range of plume shapes and sizes that may not be represented in a modelled dataset. Including modelled plumes as well as observational data in the training dataset would likely enhance the model performance but is outside the scope of this paper.

We find 1000 images (augmented to 4000 images) was sufficient to get our desired result for this project and paper. More iterations of training with more images will improve the model but with diminishing returns. As manually labelling plumes is labour intensive, we determined the 1000 would be enough to showcase the capabilities of this method.

A limitation of the machine learning model is that it cannot differentiate between merged plumes. For the model to be able to do this, it would need many (100s-1000s) examples of merged plumes which we do not currently have. With our method it would also require splitting the merged plume into two (or more) at the labelling stage which would introduce errors. As you rightly point out, another option would be to use modelled plumes, which could be used to train a model to detect and correctly identify merged plumes. This is currently outside the scope of this paper but would be a very interesting expansion to the project.

If the current model mislabelled two merged plumes as one large plume, it would not double count the emissions but calculate the total SO_2 from all merged sources. It is worth

noting that the emission estimate would likely be inaccurate as a merge plume is unlikely to be a standard plume shape (e.g. gaussian distribution in broadly one direction).