Dear reviewers and editors,

Thank you very much for your helpful and constructive feedback. In response, we have made clarifications and improvements to the manuscript, both in contents and in visual presentation, which we believe enhance the clarity and quality of the study. Detailed, point-by-point responses are shown below, with reference to the specific line numbers of tracked changes ("All-Markup" version) made in the updated manuscript.

Topic Editor's Comments on Public Justification

The reviewers and I recognize the significant improvements made to the manuscript. However, several critical issues must be resolved before the paper can proceed further.

1. Data & Methodology:
While the overall quality has enhanced, you must further address the lingering concerns regarding data representativeness, usage compliance, and methodological clarity. These are fundamental to the reliability of your results.

2. Legal Compliance & Data Sharing (Critical):
A major concern has been raised regarding your dataset's alignment with the Google Maps Platform Terms of Service. Specifically, Section 3.2.4(b) prohibits creating derivative content—including tracing, digitizing, or creating new datasets—based on Google Maps Content.

Given that your work presents a dataset intended for public research use, you must verify and provide a clear justification for its distribution. If the dataset or its labels are considered derivative products of Google Street View (GSV), their open release may violate licensing policies.

Please clarify the legal basis for sharing and reusing this data. Ensuring full compliance with third-party terms is a prerequisite for publication to avoid potential legal or policy risks for both the authors and the journal.

Response: We thank the Topic Editor for the constructive feedback. In our responses to the reviewers and in the revised manuscript, we have explicitly addressed the remaining concerns regarding data representativeness, data usage compliance, and methodological clarity.

Specifically, we discuss the role and limitations of Google Street View data with respect to spatial coverage, clarify its compliance with data usage policies, and enhance the methodological descriptions as well as the scope of the discussion. These improvements strengthen the reliability, transparency, and overall rigor of the proposed framework. We believe that these revisions substantially improve the robustness and clarity of the manuscript.

Report #1

This paper proposes a practical national-scale framework for crop type mapping based on previous research. The approach appears useful for agricultural monitoring and management. Considering that the revised manuscript builds upon methodology previously published in another journal and includes extensive uncertainty analysis, the framework appears scalable and reasonably robust. However, several issues related to data representativeness, data usage compliance, and methodological clarity should be further addressed before the manuscript can be considered for publication.

Response: We sincerely thank the reviewer for the careful evaluation and constructive comments on our work. We appreciate the reviewer's acknowledgment of the framework's potential value and scalability. Below, we provide detailed, point-by-point responses to each comment, and we have revised the manuscript accordingly to address concerns related to data representativeness, data usage compliance, and methodological clarity:

1. While the integration of Google Street View (GSV) and Sentinel imagery is interesting, and the coverage of GSV across the CONUS is relatively good, some rural roads lack GSV coverage, and many agricultural fields may be located far from major roads. Therefore, I wonder about the reliability of predictions in these areas, particularly where no GSV-based point prompts are available.

Response: We thank the reviewer for raising this important concern regarding the spatial representativeness and reliability of GSV-based point prompts. We acknowledge that Google Street View coverage is not spatially complete across the CONUS, particularly along some remote rural roads, and that certain agricultural fields may be located farther from road-accessible locations.

It is important to clarify that CropSight-US is not intended to provide exhaustive coverage of all cropland parcels. Instead, the framework is designed to generate high-quality, object-based crop type ground truth only where reliable street-level observations are available. In areas without nearby GSV coverage, crop type labels are not inferred or extrapolated based solely on satellite imagery; such fields remain unlabeled, thereby preserving the reliability of the dataset and avoiding unsupported predictions. To address the limitation of uneven GSV coverage, we adopted a stratified, spatially adaptive sampling strategy. Specifically, we select field-view GSV locations across Agricultural Statistics Districts (ASDs), within which cropping practices are relatively homogeneous, accounting for both the extent of cultivated land and the distribution of irrigation practices. This approach is intended to ensure that the resulting reference dataset and the CropSight-US dataset are representative of diverse cropping systems, climate zones, and management practices of the study site, as described in Lines 241-262 and Lines 465-494.

We also have assessed the distribution of field-to-road distances in CropSight-US. As reported in Section 5.2 CropSight-US (Lines 625-629), the field-to-road distances for sampled parcels in CropSight-US typically range from under 10 m to 150 m, and the average distance from sampled

CSB field boundaries to roads is 112.32 ± 74.66 m. These distances indicate that the majority of CropSight-US ground truth objects are located within the spatial context of typical roadside agricultural fields in the CONUS. Owing to the regular road network structure of many U.S. agricultural regions, particularly under the Public Land Survey System, a substantial proportion of cropland parcels are adjacent to or proximal to roads, enhancing the practical representativeness of roadside observations. Nevertheless, we agree that regions characterized by sparse road networks may be underrepresented, and we have included an explicit discussion on this limitation and its implications for spatial representativeness and downstream applications in the revised manuscript (Lines 750-760).

2. The study relies heavily on Google Street View data, which is subject to specific usage and redistribution restrictions. In particular, the Google Maps Platform Terms of Service (Section 3.2.4(b)) impose limitations on data storage, reuse, and redistribution. Given that this is a dataset-oriented paper, it is essential to ensure the use of GSV data complies with these terms, and that any released dataset or derived products do not violate licensing constraints.

Response: We thank the reviewer for raising this important point regarding data usage compliance and redistribution constraints associated with Google Street View (GSV). We fully acknowledge that GSV data are subject to the Google Maps Platform Terms of Service, including restrictions on data storage, reuse, and redistribution (e.g., Section 3.2.4(b)).

Importantly, the CropSight-US dataset does not include any raw or derivative GSV imagery, nor does it redistribute GSV-related visual content or metadata information that can be used to trace back to the original GSV, as we updated in the Data Availability section (Lines 788-792). All GSV images are accessed exclusively during the processing stage for the purpose of visual crop type identification, and are not stored, shared, or redistributed as part of the released dataset.

The released CropSight-US dataset consists solely of non-reversible products using GSV and Sentinel-2 imagery, including (1) object-based crop type labels, (2) field boundary geometries delineated from Sentinel-2 imagery using SAM, and (3) associated object attributes such as temporal attributes and uncertainty estimates. These outputs do not permit reconstruction of the original GSV images.

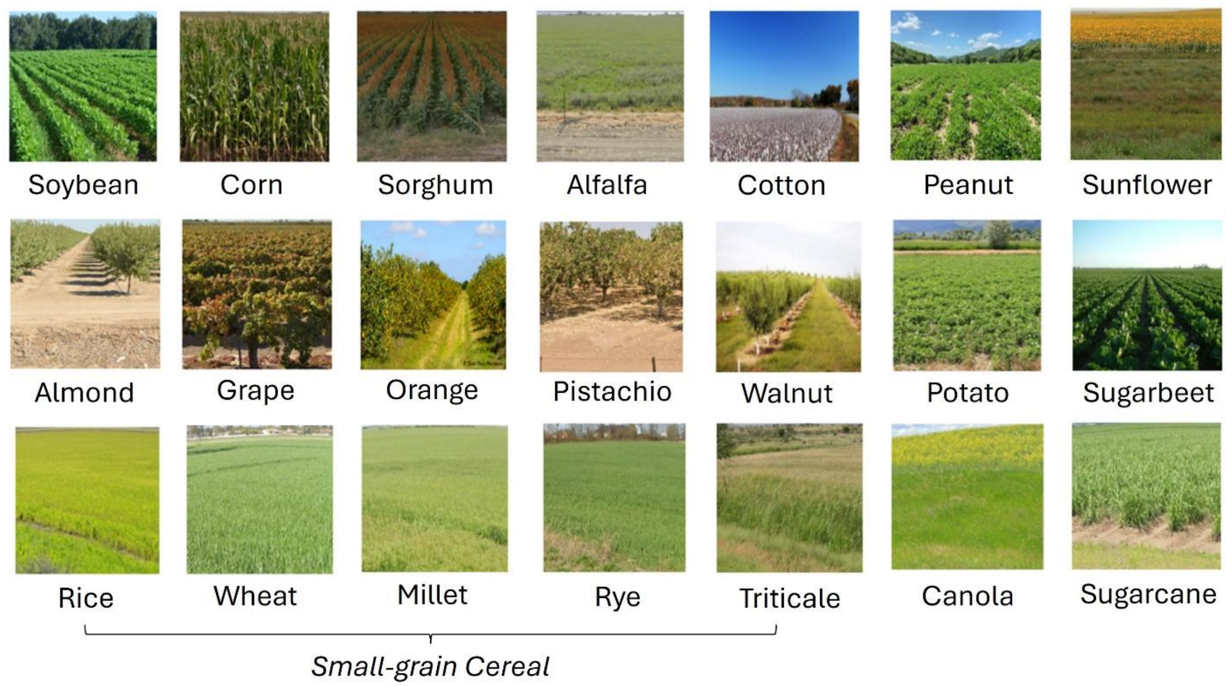
3. It would also be helpful if the authors could clarify the monetary and time costs associated with obtaining the GSV data used in this study, as this information would be valuable for assessing the feasibility of transferring the framework to other regions or countries.

Response: We thank the reviewer for this helpful suggestion. We have added a clarification of the monetary and computational costs associated with accessing GSV data in the revised manuscript (Lines 229–233). According to the Google Maps Platform pricing policy, Static Street View imagery requests are billed at USD 7 per 1,000 requests for usage up to 100,000 requests and USD 5.60 per 1,000 requests for usage between 100,001 and 500,000 requests, following a monthly free usage quota of 10,000 requests per user. Street View Metadata requests are free of charge and are not subject to request limits (Google Maps Platform Pricing: https://developers.google.com/maps/billing-and-pricing/pricing). In our implementation, the total number of image requests required to construct CropSight-US falls within the 100,001–500,000 request tier. For small-scale pilot studies or exploratory experiments, the monthly free usage tier of Google Maps Platform (10,000 requests per account) is generally sufficient to support GSV data acquisition without incurring additional costs.

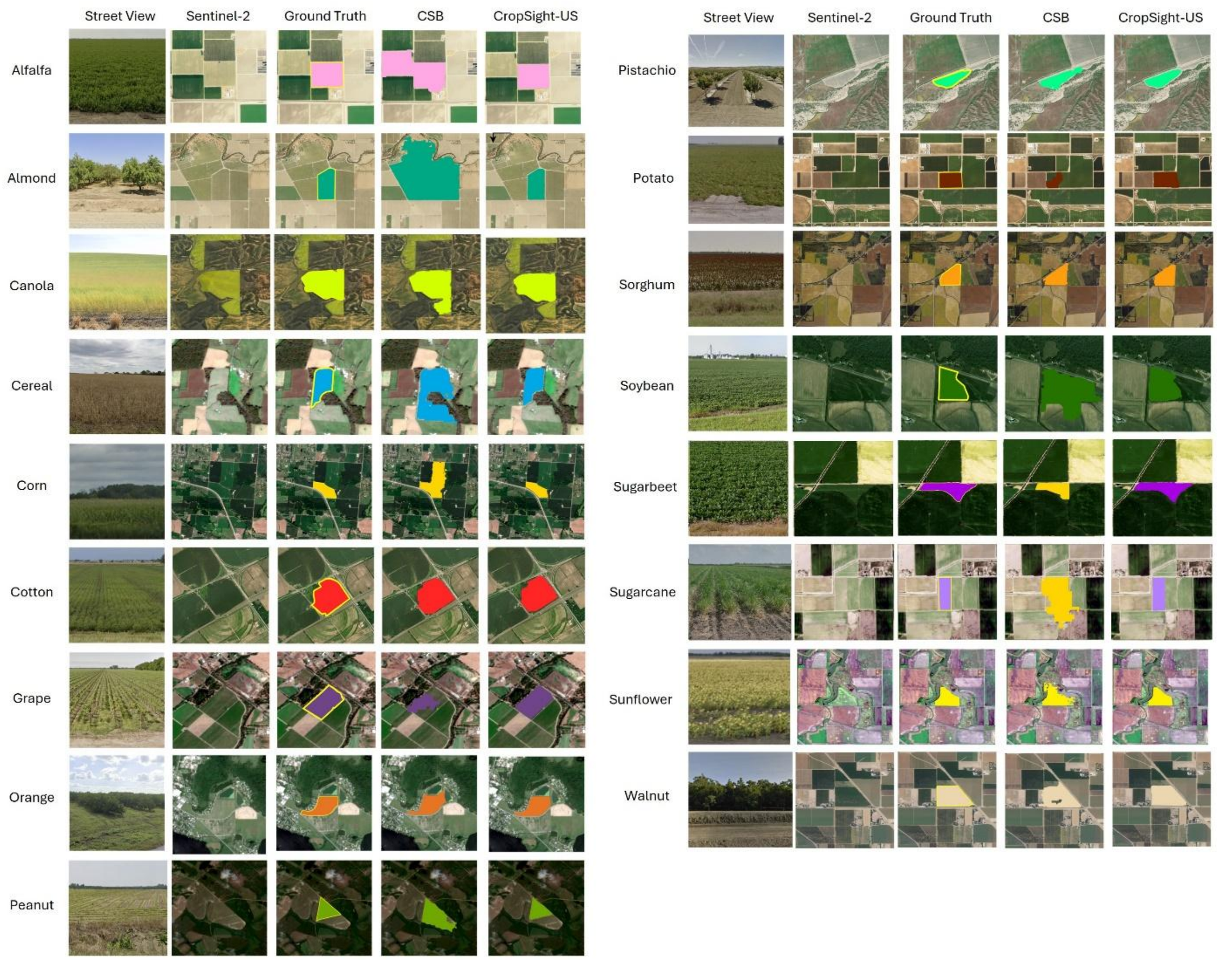
Importantly, we extensively used metadata queries to pre-filter valid and accessible locations prior to initiating image downloads, which substantially reduced unnecessary image requests and overall monetary costs. This metadata-first strategy is a key component of the framework's efficiency. In terms of time and computational costs, both GSV metadata and image requests are executed programmatically and can be efficiently parallelized. For the national-scale dataset used in this study, data acquisition required several hours to a few days of processing time on a standard workstation, depending on request throttling and batching strategies. No specialized hardware was required, although execution can be further accelerated using cloud-based or distributed computing resources. Overall, both the monetary and time costs associated with GSV data acquisition are modest, transparent, and scalable. These characteristics support the practical feasibility of transferring the proposed framework to other regions or countries where GSV coverage is available.

4. There appear to be some redundant gray lines in Figures 3 and 10.

Response: Thank you for pointing this out. The redundant gray lines in Figures 3 and 10 were likely introduced during the figure compression process when preparing the manuscript pdf for submission. We have carefully checked the original figures and corrected this issue in the revised version to ensure the visual elements are clean and accurately represented.



Updated Figure 3: Examples of the GSV component of the CropGSV-Ref reference dataset showcasing the cropland field-view GSV images of 17 crop types (GSV images are from © Google Maps).



Updated Figure 10: Examples across the 17 crop types showcasing roadside field-view GSV imagery, corresponding Sentinel-2 satellite imagery, object-based crop type ground truth, and object-based crop type information generated by Crop Sequence Boundary (CSB) and our ground-truthing framework developed to produce the CropSight-US dataset. GSV images are from © Google Maps, and Sentinel-2 imagery thumbnails are from © European Union/ESA/Copernicus, processed via Google Earth Engine.

5. The discussion section could be strengthened by considering more recent cross-view learning approaches based on deep learning.

Response: We thank the reviewer for this helpful suggestion. We agree that inclusion of recent deep learning–based cross-view learning studies could strengthen our discussion section. The framework used to retrieve the CropSight-US dataset links ground-level observations with overhead satellite imagery by associating street-level crop labels with corresponding field-scale satellite data. This linkage provides a practical use case for cross-view learning in agricultural applications and could serve as a foundation for developing and evaluating cross-view learning models. We highlight this connection in the revised Discussion section (Lines 766–770). We believe this expanded discussion better situates our contribution within the broader literature and clarifies the distinctions and complementary nature of our framework relative to recent deep learning-based cross-view learning approaches.

Report #2

The authors have addressed all concerns, and the reviewer recommends acceptance.

Response: We thank the reviewer (Report #2) for their careful evaluation of our manuscript and for recognizing the improvements made in response to earlier comments. We appreciate their support for publication.