**Original Manuscript ID:** Access-2024-41310

**Original Article Title: “**Visibility Aware In-Hand Object Pose Tracking in Videos with Transformers”

**To:** IEEE Access Editor

**Re:** Response to reviewers

Dear Editor,

Thank you for allowing a resubmission of our manuscript, with an opportunity to address the reviewers’ comments.

We are uploading (a) our point-by-point response to the comments (below) (response to reviewers, under “Author’s Response Files*”*), (b) an updated manuscript with yellow highlighting indicating changes (as “Highlighted PDF*”*), and (c) a clean updated manuscript without highlights (“Main Manuscript”*).*

Best regards,

Dinh-Cuong Hoang, et al.

**Reviewer#1, Concern # 1:** The paper are well organized and ready for publication.

**Author response:** Thank you for your positive feedback. No specific changes are needed for this concern. We will ensure all minor edits and enhancements suggested by other reviewers are incorporated to maintain the manuscript's overall clarity and quality.

**Reviewer#2, Concern # 1:** While the manuscript demonstrates improvements over existing methods, the comparison lacks depth. Consider expanding your discussion to include additional baselines or related work.

**Author response:**  Thank you for pointing this out. Our manuscript includes comparisons with several state-of-the-art methods, but we agree that a more detailed analysis could strengthen the evaluation.

**Author action:**  We will expand Table 1 to include additional recent baselines, such as [specific methods from 2023–2024], to provide a more comprehensive comparison. Additionally, we will add a subsection in the Results section to discuss the differences in performance trends across these methods in greater depth.

**Reviewer#2, Concern # 2:** Some equations and the overall algorithm are not adequately described. Ensure that each mathematical element is defined and that the algorithm is presented with enough detail for reproducible.

**Author response:**  Thank you for this valuable observation. We will revise the mathematical descriptions to ensure all variables, functions, and equations are explicitly defined.

**Author action:**

* Define all variables, constants, and parameters within the text accompanying equations.
* Add pseudocode for the main algorithm, ensuring that it is comprehensive and includes key steps like the initialization of the transformer modules and loss calculation.
* Include additional explanations in Section III (Methodology) to clarify each component’s implementation.

**Reviewer#2, Concern # 3:** The methodology omits several key details, such as specific hyperparameters used, dataset preprocessing steps, and training settings. Providing this information would enhance the clarity and reproducibility of your work.

**Author response:**  We acknowledge the importance of providing full transparency for reproducibility. We will enhance the methodology section with these missing details.

**Author action:**

* Add a new subsection titled "Implementation Details," including hyperparameters (e.g., learning rates, batch size, dropout rates), dataset preprocessing steps (e.g., image resizing, augmentation), and training environment settings (e.g., GPU configuration).
* Provide detailed descriptions of optimizer settings and learning rate schedules.

**Reviewer#2, Concern # 4:** Some figures and plots lack sufficient annotations or context. Enhance these visual aids to ensure they are self-explanatory and clearly support the narrative.

**Author response:**  Thank you for pointing this out. We will enhance the clarity of figures and plots to ensure they are fully self-explanatory.

**Author action:**

* Add detailed captions for all figures, explicitly describing axes, legends, and any highlighted features.
* Enhance Figures 2 and 3 by including zoomed-in views for occluded scenarios to emphasize differences in performance.
* Annotate key trends and outliers directly on performance graphs in the Results section.

**Reviewer#2, Concern # 5:** Discuss the computational demands of your approach, particularly with regard to the transformer architecture, and propose strategies for real-time implementation.

**Author response:** We appreciate the suggestion to include a discussion on computational efficiency. Addressing this concern will highlight the practical applicability of our approach.

**Author action:**

* Add a new subsection in Section IV (Evaluation) discussing the computational overhead of transformers, including a breakdown of inference time and memory usage.
* Propose strategies such as model pruning, quantization, and the use of lightweight transformer variants (e.g., MobileViT) for real-time implementation.
* Include comparisons with faster, albeit less accurate, methods to demonstrate the trade-offs.

**Reviewer#2, Concern # 6:** Provide a clearer distinction between your visibility-aware module and similar mechanisms in the literature to highlight the unique contributions of your work.

**Author response:** Thank you for this important suggestion. We will refine the manuscript to better highlight the unique aspects of our visibility-aware module.

***Author action:***

* Expand the Related Work section to compare our visibility-aware module with prior occlusion-handling approaches.
* Clarify the key innovations, such as the use of temporal pose aggregation via cross-attention, that distinguish our module.
* Include a dedicated paragraph in Section III to explicitly describe these distinctions.

**Reviewer#2, Concern # 7:** While the references are generally relevant and up-to-date, consider including additional recent works that directly address visibility-aware tracking or transformer-based object pose estimation to strengthen the contextual foundation.

**Author response:**

- We agree with your suggestion and will update the references to include more recent works that address visibility-aware tracking and transformer-based pose estimation.

***Author action:***

* Incorporate 5–6 recent references (2023–2024) into the Related Work section.
* Highlight these works explicitly in the discussion to provide a stronger contextual foundation for our contributions.

**Reviewer#2, Concern # 8:** A section discussing the limitations of your approach, such as scalability to larger datasets or performance in real-time scenarios, would provide a more balanced and transparent evaluation of your work.

**Author response:** We appreciate your suggestion to discuss the limitations of our approach, as this adds balance and transparency to the evaluation.

***Author action:*** Add a dedicated "Limitations and Future Work" section discussing scalability challenges, the trade-off between model complexity and speed, and potential issues in highly cluttered environments.

**Reviewer#2, Concern # 9:** While the manuscript is written in readable English, there are minor grammatical errors and phrasing inconsistencies. A thorough language review is recommended.

**Author response:** We will conduct a thorough language review to correct any inconsistencies.

***Author action:***