We thank the Associate Editor and both reviewers for their constructive feedback on our manuscript. We appreciate the reviewers’ positive assessment of our work and their valuable suggestions for improvement. We have carefully addressed all comments in detail below in point-by-point responses to each reviewer comment, with changes to the manuscript highlighted in yellow.

**Reviewer 1**

RE1.1:  
**Keywords: The first used keywords are already part of the title. To improve SEO, I suggest replacing them with further information about the sequence (i.e. trajectory?) or algorithms used in segmentation**

Thank you for this suggestion to improve SEO. We have replaced the previous keywords

”Dynamic MRI; Bone tracking; Semi-automated segmentation”

with these: Radial golden-angle acquisition; Canny edge detection; CINE reconstruction; Knee osteokinematics

Materials and Methods

RE1.2: **I would prefer a photograph of a subject in the motion device. I don't find the pictures in [25] to be very helpful. I suggest an image where the patient table is in its home position, to better show the deflection of the lower leg.**

Thank you for this thoughtful suggestion. While we agree, taking a photograph of the device within the MRI scanner from a closer distance and different angle (to better show the deflection of the leg) is difficult due to the magnetic field. The images provided in [25] are what we could obtain without jeopardizing participant safety. However, [25] also includes a video as supplementary material in the online version of the publication. This video shows the leg and motion with the table in the home position. The direct link to the video:  
<https://www.sciencedirect.com/science/article/pii/S093938892100115X?via%3Dihub#upi0005>

RE1.3:  
**To estimate the flexion-speed better, I suggest to report the cycle time, rather than the cycles per minute. This way it is easier to relate motion speed and sequence temporal resolution.**

This is indeed a good suggestion for improving interpretation. Accordingly, we have revised the text to report cycle time instead of cycles per minute. The updated text, shown on Page 5, now reads: “Each knee extension-flexion movement cycle was guided by eight metronome beats, with the knee fully flexed at the first beat, fully extended by the fourth beat and fully flexed again by the eighth beat, resulting in 8 seconds per cycle”

RE1.4:  
**TGV was used as a regularizer. Along which dimensions? time, space, or both? Please clarify.**

The TGV regularization was applied along the spatial dimension. The updated text, shown on Page 6, now reads: “... the image was reconstructed as a spatially regularized Total Generalized Variation ~~regularized~~ least-squares problem solved using the Alternating Direction Method of Multipliers, ...”

RE1.5:  
**Is there a python image-processing package that was mainly used for the analysis (i.e. openCV)? If so, please mention the used package.**

Indeed, two image processing packages were used for the analysis: SciPy and sciki-image. We have now mentioned the specific modules used for edge detection, connected-component labeling, cubic spline interpolation and Nelder-Mead optimization alongside the corresponding algorithms on Pages 6 and 7.

RE1.6:  
**In (I) you write: "…including the interior cortical bone boundaries." I suggest: "including the boundary between cortical and trabecular bone"**

Thank you for this suggestion. We have revised the text accordingly on Page 7, as shown here: “This step resulted in binary images highlighting the detected edges, including the boundary between ~~interior~~ cortical and trabecular bone.”

RE1.7:  
**In (III) you write that the most distal points were used as initial points. Would it not be the most proximal point for the tibia?**

We used the most distal point for both the femur and tibia for simplicity in algorithm implementation. However, the choice of the starting point (distal vs. proximal) does not affect the tracking performance, as the algorithm establishes reference points along the entire bone contour.

RE1.8:  
**What exactly justifies the 'semi' in the presented technique? I assume it is the selection of the connected component, that represents the bone (Step II)? Please clarify.**

Thank you for seeking clarification. Indeed, the selection of connected component requires manual intervention, as does the tuning of edge detection parameters. To clarify this point in the manuscript, the following text has been added on Page 8: “The semi-automated approach required manual intervention at two stages: one-time optimization of edge detection parameters for the given image contrast and resolution, and manual selection of labeled components representing the bone edges of interest in the reference frame, performed once per dataset.”

Results

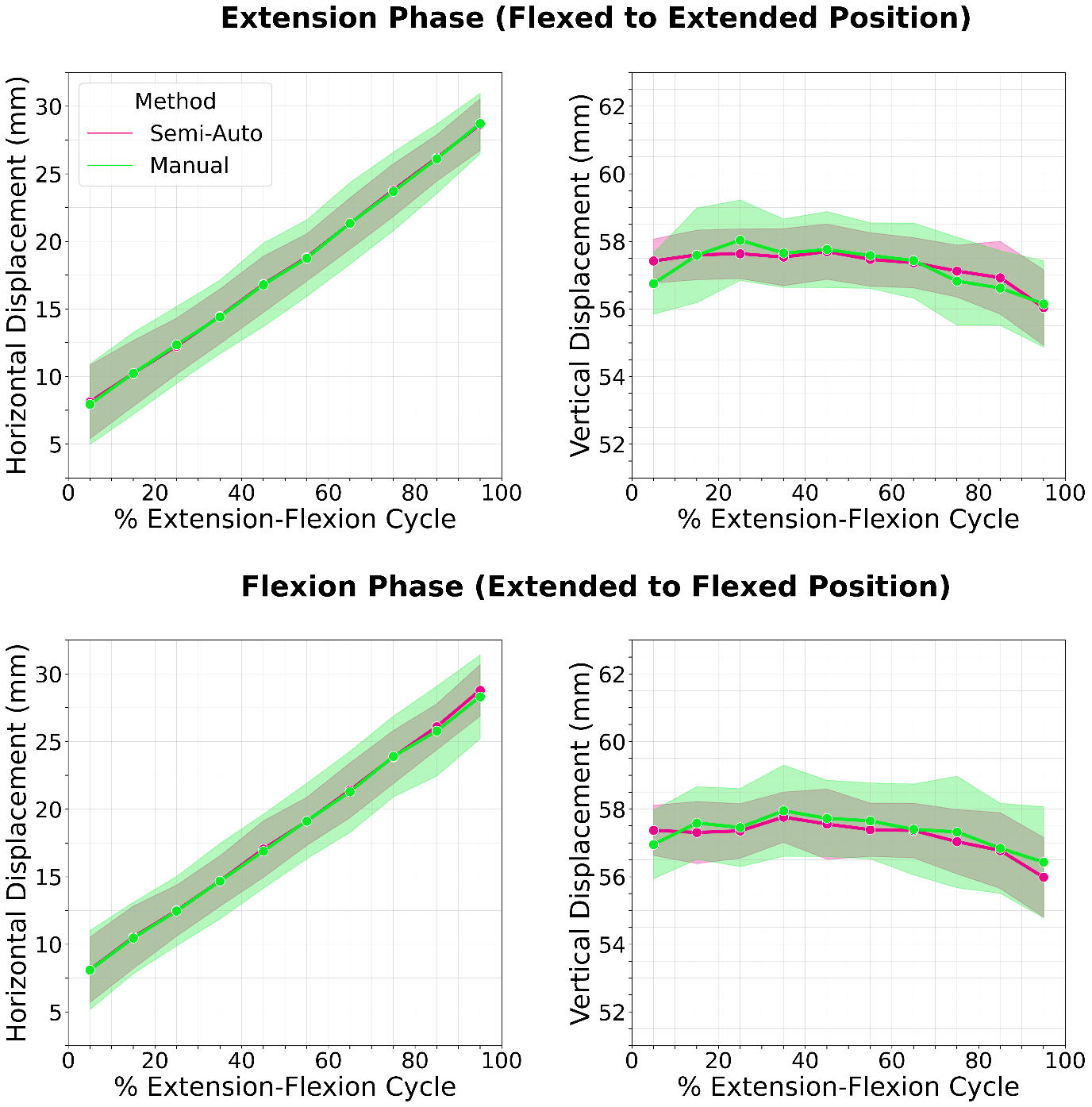
RE1.9:  
**You mention the duration of both, manual and semi-automatic segmentation approaches. On what kind of machine was this measured?**

Processing times were estimated during the development and validation of the method. For manual segmentation, processing time was estimated using a stopwatch during the process. For the semi-automated approach, processing time includes both the computation steps (edge detection, transformation optimization) and manual interventions (parameter selection, component labelling), and was estimated on a standard desktop workstation with Python (start to end of the processing script). The reported processing times represent typical durations observed across the datasets. The following information was added to the manuscript on Page 9:

“For manual segmentation, processing time was estimated using a stopwatch; for semi-automated segmentation, processing time was measured on a standard desktop workstation running Python and included both computation steps (edge detection and transformation optimization) and manual interventions (parameter selection and component labelling).”

RE1.10:  
**Figures 3 and 4 use the same colors for different things (femur/tibia and semi-auto/manual) please use different colors to avoid confusion.**

Thank you for this suggestion. The colors in Figure 4 have been modified as shown below:



RE1.11:  
**I am unsure how the data points in Fig4 are generated. From Fig3 I assume there is about 30 frames for a full flexion cycle. Fig4 only shows 10 for a half-cycle. How exactly are data from different frames (and subjects) combined?**

Thank you for seeking clarification. Regarding Figure 3, 8 frames out of a total of 28 frames (index 0 to 27) are shown. Regarding Figure 4, different subjects achieved different ranges of knee motion, resulting in different numbers of frames. To deal with this, we binned the normalized flexion cycle into 10% intervals and averaged displacement values within each bin across all datasets.

To clarify these points in the manuscript, the following text has been added on Page 9 (Section 2.3): “Due to inter-participant variations in frame counts resulting from these motion variations, the normalized flexion cycle data were binned into 10% intervals (i.e., 0-10%, 10-20%, etc.), with displacement values averaged within each bin across all datasets.”

RE1.12:  
**Fig4: In the Horizontal displacement plots, one vertical tick/gridline represents a step of 1,25. This seems unusual and can be easily mistaken. I suggest using 1mm or 2.5mm as a tick size.**

Thank you for this suggestion. The minor tick size has been changed to 2.5 mm, as shown in RE1.10 above.

RE1.13:  
**I would like to see the following as a supplementary material: A video of the CINE images of a full flexion at similar speed to the actual knee motion. Followed by a second full flexion, now with the segmentation overlays as in fig 3 and indication of estimated centroid positions. One subject is sufficient.**

It is our pleasure to provide the requested videos ( provided in the supplementary section). The following sentence has been added on Page 8:   
“Videos demonstrating the complete CINE image sequence at physiological speed and the semi-automated tracking results with centroid positions are available in the Supplementary materials section.”

**Reviewer 2**

RE2.1  
**The knee joint also includes the patella (or kneecap), the movement of which can be abnormal in some diseases or after injuries. Why was the patella, which is well delineated in the sagittal images, not included? Would there be more problems here?**

Thank you for this important question. We initially considered including the patella in our analysis. However, the patella presented two significant challenges: first, during edge detection, the patella produced inconsistent and fragmented boundaries due to its smaller size and variable contrast with surrounding tissues, resulting in less reliable feature extraction compared to the tibia and femur. Second, the patella undergoes significant through-plane movement during knee flexion-extension, making it incompatible with our 2D rigid-body tracking approach that was optimized for the predominantly in-plane motion of the tibia and femur.

The following sentence has been included in the Discussion section on Page 12:   
“Due to this limitation, the patella was also excluded from our analysis despite its visibility in the sagittal images, as it undergoes significant through-plane motion during knee flexion and extension that is incompatible with our 2D tracking approach.”

RE2.2  
**The range of motion during flexion in the knee joint in the closed MRI unit with relatively small bore (3T Siemens Prisma) is limited depending on the length of the lower leg. It should be pointed out in the paper that full-range examinations are only possible on open MRI systems, on which the proposed methods should also work.**

Thank you for highlighting this important point. As correctly noted by the Reviewer, our method is indeed constrained by the knee range of motion achievable in closed-bore MRI systems, and open-bore systems would allow for full-range examinations. Since our edge-based tracking operates on 2D sagittal plane images, the method should be directly transferable to open-bore systems. We have added the following text to the Discussion section on Page 12:

“Furthermore, the knee range of motion achievable in closed-bore MRI systems is limited by the bore diameter and the length of the lower leg, precluding the assessment of deep/full knee flexion. In the current study, the bore diameter was 60 cm and the lower leg of the participant ranged from XX-YY cm, allowing for knee flexion angles between 30° and 46°. Open-bore MRI systems would be better suited to perform full knee range of motion examinations [42], on which the methods presented in this report should also work.”

RE2.3  
**In some orthopaedic examinations (e.g. following cruciate ligament ruptures), movements in the knee joint are measured when force is applied (e.g. anterior and posterior drawer test). Could the method also be used for quantitative assessment of such experiments?**

Thank you for this thoughtful question. Indeed, our dynamic MRI setup and centroid-based 2D bone tracking algorithm could theoretically be used to quantify knee joint laxity/stability, for instance, after cruciate ligament injury, by measuring the antero-posterior movement between the tibia and femur. Such quantitative assessments could provide insights similar or complementary to that obtained from traditional orthopaedic examinations focused on sagittal (single plane) tibiofemoral translations such as the anterior-posterior drawer test and the Lachman test. While these tests apply passive linear anterior-posterior forces to the knee joint, our method would apply rotational forces (joint moments) during volitional flexion and extension, which may in fact provide insight that holds greater clinical and physiological relevance due to the biomechanical, muscle-driven nature of the joint movement.

We have added this the following passage on Page XX to convey this notion:

“XXXXXXXXXXXXXXXX.”

RE2.4  
**How can an 'alignment error' be determined that is significantly smaller than the spatial resolution of the measurement sequences?**

Thank you for raising this point. The alignment error can achieve sub-voxel precision primarily due to the cubic-spline interpolation step, where the initially discrete edge pixels are converted into 80 continuous reference points along the bone boundary. Subsequently, the transformation matrix and distance measurements also remain in continuous space. Thus, the reported alignment error (0.40 ± 0.02 mm) reflects a continuous spatial registration accuracy metric rather than implying any resolution of anatomical details smaller than voxel spacing (1.09 mm). The following text has been added to Page X (Section 2.2 in Step III):   
“This interpolation process converts the initially discrete edge pixels into continuous coordinate reference points, enabling sub-voxel precision in subsequent transformation and alignment calculations.”

RE2.5  
**The recording method with 2D radial GRE sequences has not become completely clear. It is described that the slice thickness is 1 mm, but the FoV is 3 mm thick. How many (sagittal) slices are acquired?**

Thank you for noticing this mistake. It is a single 3 mm thick sagittal slice acquisition, but we had mistakenly written an incorrect voxel size of [1.09 x 1.09 x 1] mm3. It has now been corrected to [1.09×1.09×3] mm3 on Page 5 of the manuscript, as shown below:

“MRI data were acquired using a 2D radial golden-angle gradient echo FLASH sequence [26,27] with the following parameters: echo time of 2.51 ms, flip angle of 8°, field of view of [192×192×3] mm3, matrix size of [176×176×1], voxel size of [1.09×1.09×3] mm3, and repetition time of 5.8 ms. This acquisition protocol captured a single 3 mm thick sagittal slice.”

RE2.6  
**The shape of the bones on the images changes if the lower leg has motion components perpendicular to the slice or rotational components. In this case there are no longer matching reference points on a fixed sagittal slice. How is this handled? Should layers be reconstructed from 3D data sets that depict the same sagittal plane through the lower leg at different knee flexion angles?**

Thank you for raising this important point. Through-plane motion is indeed a limitation of our 2D tracking method, which could only be overcome using 3D methods. We have clarified this point on Page 12, as shown below:

“Despite these advantages, a limitation of the current 2D approach is sensitivity to through-plane motion. While the knee motion/loading device was designed to constrain movement to the sagittal plane, this remains a potential source of error as physiological knee flexion and extension is not purely a planar motion. This limitation also led us to exclude the patella from our analysis despite its visibility in the sagittal images, as it undergoes significant through-plane motion during flexion-extension that is incompatible with our 2D tracking approach. In cases where significant through-plane motion occurs, the bone appearances in the fixed sagittal slice change, resulting in elevated cost function values that indicate compromised tracking accuracy. Future work could address this limitation by extending the method to 3D acquisitions.”

RE2.7  
**A "semi-automated pipeline" is reported. It should be indicated at which points interventions by the examiner are necessary.**

Thanks for seeking clarity. We have now clarified on Page 8 at which points intervention by the examiner are necessary for the semi-automated pipeline, as shown below:

“The semi-automated approach required manual intervention at two stages: one-time optimization of edge detection parameters for the given image contrast and resolution, and manual selection of labeled components representing the bone edges of interest in the reference frame (once per dataset).”