# Author Response to Reviewers for the paper 'Eliminating prior-bias from sparse-projection tomographic reconstructions'

July 22, 2020

**Organization of this response:** We would like to thank the editor and all three reviewers for their insightful comments. Here, we are including our response to each reviewer, and pointers to modifications (suggested by the reviewers and the AE) in our revised manuscript. Where necessary, we tie up related comments from the multiple reviewers for ease of understanding. We have organized this document in three parts (a) Response to comments from AE (Section A), and (b) Response to significant points made by all the other reviewers (c) Detailed responses to the comments of each of the reviewers.

To the best of our knowledge, we have painstakingly addressed each and every point of all the reviewers, and responded adequately. We request the reader's time to go through our responses.

In some cases, we provide results in this document to justify our stand, but these same results are not in the updated manuscript, primarily because these results do not enhance the claims in our paper . In most other cases, we have shifted such results to supplementary material and have referred to them in this document.

For ease of understanding, since the current version of the manuscript is considerably reorganized, in this document, we have used the (longer) term "Figure x" to refer to a figure in the old version and "Fig. x" to refer to the related figure in the current version. Similarly, we have used the term "Equation y" to refer to an equation in the old version, and "Eq. y" to refer to an equation in the new version, and "Table y" to refer to a table in the old version and "Tab. y" to refer to a table in the new version.

## 1 Associate Editor comments

(We have numbered the original review comments as AE-Cx, and italicized for clarity)

Two of the reviewers have raised some serious concerns about the paper, including one reviewer who opines that this work is not suited for publication in the transactions in its current form. While these concerns are sufficiently serious

to merit a negative decision, I have decided to give the authors an opportunity to revise the manuscript so that they may address the constructive comments of the reviewers carefully and thoroughly followed by another round of review to assess the manuscript. In addition to the reviewer specific comments, please pay specific attention to the following:

**AE-C1.** *Overall organization of the paper to improve the exposition.*

**Response:** We have now made the following significant changes in the organization of the paper: The introduction has been modified and the stage set for outlining our contributions have been made. We have replaced the earlier Figure 1 (which did not have longitudinal data) with a more appropriate figure (Fig.1). The unfamiliar terms "unselective" and "selective" prior have now been replaced by "uniform" prior and "spatially varying" prior to better communicate the property of the respective priors. A new section (Section 2) named "Related Work" has been included separately to contrast this work with the prevailing work in the literature. Details of the datasets have been moved to a separate section (Section 4) so the ground is properly set. The two earlier sections: methods and results sections have now been split into the following 4 sections. This ensures better readability and emphasizes the cases where both uniform prior and spatially-varying prior are useful.

- Section 6: Uniform prior-based reconstruction
- Section 7: Results: Reconstruction by uniform prior
- Section 8: Spatially varying prior-based reconstruction
- Section 9: Results: Reconstruction by spatially varying prior

For greater clarity, we have now removed Schematic 1 (as suggested by R1) from the paper. We have ensured that the rest of the paper can be followed without it. We have also moved the discussion on the earlier patch-based methods to supplementary material. However, results not reported earlier prior to this submission using the uniform prior on 3D volumetric data continues to be present (as mentioned above) in Section 6 since it forms an important part of our story for this paper.

**AE-C2.** *Highlight clearly the novelty of the presented work and provide accurate comparisons to relevant algorithms in the case when one has access to prior scans.*

**Response:** We have now re-written the contributions section (Sec.3 of the main paper), that highlights the novelty of our work. As a response to the comparisons required, in many cases, our methods are incomparable to previous algorithms because of the nature of the longitudinal study We discuss this in the related work section. Further we reiterate that in our earlier manuscript version, we had compared our method with l1-regularized least squares (l1-ls) and FDK for 3D reconstructions. We had also compared (Fig.8) our method with many other iterative methods such as (ART, SART and SIRT) – this set of

comparisons was buried in a footnote, and perhaps not easily visible. In addition to these, we now have included Table 3 to show a quantitative comparison of SSIM values of the reconstruction by various methods shown in Fig.8. We also discuss compressed sensing (see AE-C3) results. In addition, R3 has referred to "Region-based iterative reconstruction of Structurally Changing Objects in CT". This reference is now mentioned in our related work; although this paper is not directly applicable in longitudinal studies where object-priors 'are' available, there is a reference in this paper to another paper which does make use of object-prior. We show how our work is superior in our response to R3 below (we include this in the supplementary material).

**AE-C3.** *Discuss the compressed sensing methods used since there are several works on model-based iterative reconstruction/regularized inversion/CS type methods for CT.*

**Response:** We believe we now have clearly brought out the compressed sensing based works both in the introduction and in the related work section. Related points w.r.t. CS-type reconstruction made by R1/R3 appears below in B1 (and elsewhere in even greater detail).

In summary, we believe that the extensive work we have done based on the reviewer comments, while still maintaining the intellectual contributions of the original submitted version lends itself to the paper being accepted.

## 2 Broad response to all reviewers

We believe it is useful to single out the most serious criticism to this paper, and we wish to defend our stand. At this time, we request the reviewers to kindly look at the first few sections, with slightly modified terminology (compared to the submitted version) before continuing on to the responses.

- R1 and R3 are concerned about the blurring and over-regularized nature of the compressed sensing-based images. In particular R3 is seriously concerned about the quality of Compressed Sensing reconstruction. Our response is that this is a function of the number of views. In short, if the number of views is sub-Nyquist, yes, indeed CS can provide good quality reconstruction, especially with the tuning method suggested by R1. (Our updated figures demonstrate this). However, coming to the main contribution of the paper, if the number of views becomes significantly small (when compared even to sub-Nyquist), reconstructions using CS is less than satisfactory, despite the removal of sub-sampling artefacts. This has also been pointed out earlier in a different (not ours) published work [11], Page 4, line 22. We have now described the use-case for CS reconstructions in Fig. 1 and have elaborated on the blurred artefacts in our response in R3-C7.

- All reviewers are concerned about the seemingly low SSIM value (relatively poor reconstruction) for one of the experiments. (Figure 10-c in the

original document) while results in all the other experiments are acceptable to prove the point. For instance R2 says "I have one main question concerning Table 1, where I notice FDK gives for the ROI much better SSIM value than the variants of the proposed method." Of course, we never intended to *not* mention this, which is why it was picked up very easily by all the reviewers, and in real life situations, it is unreasonable to expect that the ROI of detailed changes is known in advance.

In response, we first note that our updated results of comparison for this dataset (after the tuning suggested by R1) *now* results in a favourable result (Table 3). That said, we agree that the FDK reconstructions can be superior, as evidenced for this particular dataset alone – presumably when the region of interest is known a-priori (which is not always the case). We have devoted a separate section in the 'Discussion' for this particular concern: Tab.8 better explains why our method is suitable, and Fig. 19(e) the more pleasing visual result.

- R1 and R3 also raise another significant point: What happens if we have only the object prior term and ignore the sparsity prior?. This indeed is relevant. Our response is that the answer to the question revolves around the quality of the prior. In our experiments, the object priors are of high quality, the number of views very few (the title of our paper), and hence the object-prior has a dominant effect when compared to sparsity prior; adding the sparsity prior improves the reconstruction marginally. That said, ignoring the sparsity prior would be akin to ignoring the established work revolving around compressed sensing. If the object-priors were of poor quality, or if the number of views (still sub-Nyquist) were to be increased from our very few views, then the sparsity term will start playing a part.

- R3 remarks that 'it's simply not true' about some of the overarching broad statements which we had claimed in our earlier submission (for example, "in most of the literature on tomographic reconstruction, the results are shown on reconstruction from projections simulated from 3D volumes"). In our current version, we have now either given better context to such statements or have retracted our misunderstandings with the wisdom in R3's remarks.

# 3 Detailed response to individual review comments

## 3.1 Response to Reviewer 1

There are some concerns associated with the compressive sensing (CS) method, the explanation for weight selection, and the clarity of results. Some of the major issues are described below.

**R1-C1:** *In Fig. 4, the caption says CS and no prior. This is not true. CS does have a prior in the sense that it forces the reconstruction to lie in a sparse basis. You did explain this aspect of CS in section III (A).*

**Response:** Agreed. CS does use the sparsity prior. We have now moved this point much earlier in the introduction, and point out some of the advantages of using object-prior. Sec.1, Page1. We have corrected the captions. Specifically, (for example) the earlier caption of Fig.4-c: "CS and no prior resulting in blurred bone structures" has now been replaced by "l1-ls resulting in blurred bone structures". In the latter, "l1-ls" refers to L1-regularized least squares optimization

**R1-C2:** $\lambda_1$ *seems to decide the strength of the CS prior. However, in section V (B), you mention that* $\lambda_1$ *is always chosen to be one. But, it is clear that the potato image in Fig. 10 (c) is over-regularized. It is extremely blurred.*

**Response:** Agreed. See response is **R1-C3**.

**R1-C3:** *In CS, since you use cross-validation on a test image to select* $\lambda_1$*, it is necessary to tune this parameter until you reach a desired SSIM or root-mean-squared-error (RMSE). With CS priors, I would expect increasing* $\lambda_1$ *to also lead to smoother results.*

**Response:** With respect to Figure. 10(c) of the submitted paper which corresponds to Fig.17(c) of the new version, we have now updated the reconstruction after tuning $\lambda_1$ to get the optimal result for l1-ls, as suggested by Reviewer 1. The optimal $\lambda_1$ was found to be 1.7.

In Sec.5 of our supplementary material, we describe why the image in Fig.17(c) (Fig.10(c) of the old version) appears over-regularized by presenting reconstructions on 2D potato dataset for different number of views and optimally tuned $\lambda_1$ values.

**R1-C4:** *What is the sparsity basis for CS prior in equation (1)?*

**Response:** We have chosen DCT as the sparsity basis, and have now clarified this in the paper. We had observed that there is no significant improvement if, say, the wavelet transform were instead used as the basis.

**R1-C5:** *Why is a CS prior necessary when you have the selective/unselective prior? What happens if there is no CS prior and if we use only the selective/unselective priors?*

**Response:** Although the data-agnostic CS/sparsity prior and the data-dependent object-prior (either uniform or spatially varying) affect independent terms in the overall cost function (Eq.6), best results are obtained when both the priors are used together. Whether to use the CS prior or not is a function of the quality of the object-prior. In our case, the object priors are of high quality, and hence the object-prior has a dominant effect. Adding sparsity prior improves the reconstruction marginally.

**R1-C6:***Schematic 1 in section III (B) does not make sense. But, the explanation in section III (B) (2) is clear.*

**Response:** For greater clarity, we have now removed Schematic 1 from the paper; We have ensured that the rest of the paper can be followed without it.

**R1-C6(b):***found section III (B) (2) algorithm to compute weights map W very interesting. However, it is hard to understand what artifacts can be attributed to the system geometry and an algorithm. For instance, in the third step, you mention that the difference between $P^{fdk}$ and $X^{fdk}$ will not contain false positives due to imaging geometry but will contain false positives specific to the reconstruction method. Can you show these artifacts for the potato dataset? While Fig. 7 showing weights is interesting, from a clarity standpoint, I would like to see images of artifacts that get removed at each stage of your algorithm (2) Algorithm to compute weights maps W.*

**Response:** To illustrate how different reconstruction methods lead to different artefacts and how combining them carefully can cancel those artefacts, we present a few intermediate weights maps for a 2D reconstruction from 12 views (of the test image in Fig.10 of the paper), with and without noise. Due to the lack of space, we have not included the below images in the main paper.

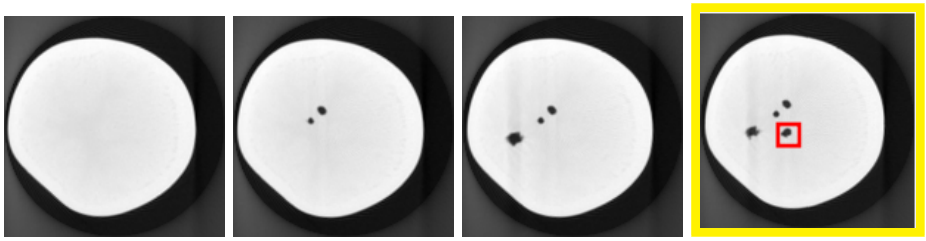

Figure 1: Potato dataset: One slice each from the previously scanned objects (the first three from left) and a slice from the test volume (extreme right). Notice the appearance of the fourth hole in the test slice.

In the Figs. 2 and 3, notice how the weights map is different for each of the methods and how it is improved when it is computed using two or more methods. We observe similar benefits even in the presence of small quantities of noise as shown below.

In the actual reconstruction of the images, the prior term is well represented and these effects are not immediately visible. However, with low quality priors the effects may be visible.

**Case 1: No noise in projections:**

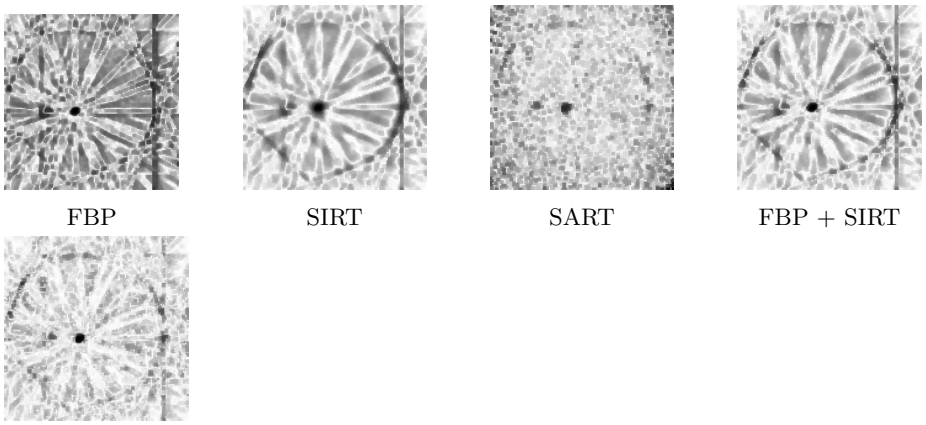


Figure 2: Weights images by various methods while reconstructing the test in Fig. 1 from 12 views and no noise

**Case 2: Noise with 0 mean and SD = 1% of the mean of the measurements**

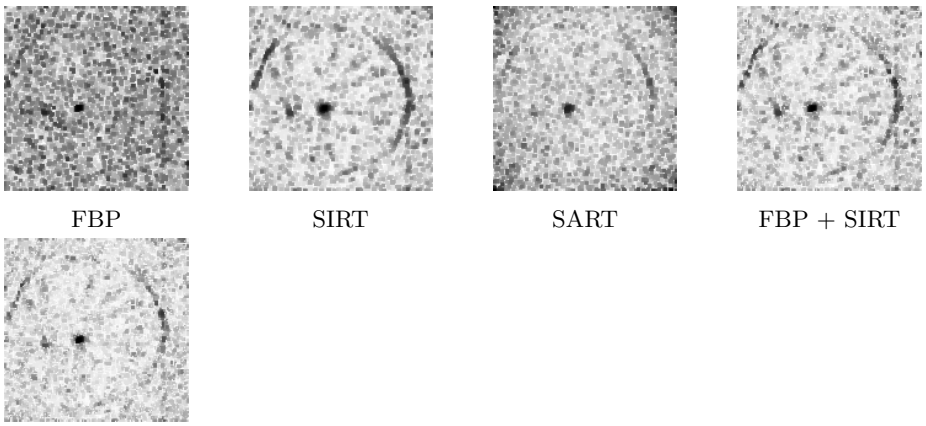


Figure 3: Weights images by various methods while reconstructing the test in Fig. 1 from 12 views in the presence of noise.

**R1-C6:** *In results, the images are simply too big relative to the ROIs, which are tiny. It is very difficult to see the changes within the ROI. If the paper is printed, then there is no way to zoom into the images. It will be better if you*

*could also present zoomed images of the ROI. Otherwise, you could present one full sized image and rest of the images could be only the ROI.*

**Response:** While the zoomed-in RoI of Okra is in Figure 13 of the old manuscript (Fig.14 of the new manuscript), the zoomed-in RoI of the sprouts dataset has now been added in Fig. 16. For other cases, larged sized images are in the supplementary material.

**R1-C7:** *Table I, II, and III present SSIM comparison. It will also be interesting to present a root mean squared error (RMSE) comparison. RMSE measure will help in evaluating the quantitative accuracy of your reconstruction.*

**Response:** The optimal values of parameters obtained by tuning RMSE were found to be different from those obtained by tuning SSIM. Hence, for consistency, we choose only one metric (SSIM) to report accuracy of all reconstructions. Moreover, SSIM gives us the flexibility to measure the fidelity of the structural content of our reconstruction with respect to that of test volume.

**R1-C8:** *For instance, in Fig. 10 (e), the region within the ROI seems to be brighter than the surrounding holes. This may also happen if an incorrect choice of the CS prior regularization parameter $\lambda_1$ results in excessive blurring.*

**Response:** Despite tuning $\lambda_1$, In Fig. 19(e) (previously Fig.10(e) of the old version), the information about the new region (within RoI) is present only in the measurements of the test, and is absent in all of the templates. This is unlike the case for the surrounding holes, the information about which is present both in the measurements of the test and the templates. Hence, the reconstruction of the hole in the RoI is not as sharp as the other holes.

**R1-C9:***Some relatively minor issues are below- Page 3, line 7, section II: Consider defining longitudinal study.*

**Response:** We have now defined longitudinal study in the paper (Sec. 1). Longitudinal studies here refers to the acquisition of sequential CT scans of the same subject to track time-evolving changes within the subject's interior.

**R1-C10:** *Images in Fig. 3 seem to be clipped in the lower left along a circular arc! Are all reconstructions within a circular outer region of interest?*

**Response:** All the reconstructions are within the circular region for this dataset. The ground-truth for this dataset consists of reconstructions from a proprietary software used in the CT machine at the Tata Memorial Hospital, Mumbai. The RoI was chosen by the proprietary software and hence the exact reason for its choice is known only to the vendors. Based on inputs from physicians at the Centre, we have ensured that the RoI indeed encompasses the regions of interest, i.e. the organ in which ablation is being performed.

**R1-C11:***In Fig. 8, what is (g) combined? Also, there is no (h) but is mentioned in caption.*

**Response:** The caption has now been correctly renamed to (g). By 'combined', we denote the weight map computed as per Eq.7. This equation incor-

porates the use of the eigen-space of the templates reconstructed from various methods (such as ART, SART, SIRT, etc.), and pilot reconstructions of the test, again reconstructed from various methods. In Eq.7, the min(.) operator ensures that we select the new regions that are unanimously detected by all the methods of reconstruction. Hence, the word 'combined' denotes that all the reconstruction methods are combined to give information about the new regions of the test.

**R1-C12:** *Try to mention the actual number of views/projections instead as a percentage. The percentage only makes sense when you are comparing against Nyquist limit for views. For instance, the total number of views for Potato dataset is 900, which is simply too many for a reconstruction size of 150x150. For such a small reconstruction, 900 is not a "typical" number views. It may be more "typical" to use 150 views in standard practice. So, expressing views as a percentage of 900 is confusing.*

**Response:** Just to clarify, as mentioned in the paper, the 900 views for the potato dataset was used to reconstruct the ground-truth 3D volume of size 150 x 150 x 100, and not a 2D slice of size 150 x 150. Each of the projection images was of size 150 x 150.

Since our imaging method is circular cone beam projection, let us first assume a 2D example. Consider a circle of radius R=1 that envelops an unknown 2D region X. Then, the circumference of the circle in pixel units will be $2\pi R_p$, where$R_p$ represents the radius in pixel units, i.e., the number of image-pixels that fit within the radius of the circle. Now, let us assume that we want to pass rays around a half of a circle (due to symmetry). We can reasonably assume that a set of $(2 * \pi * R_p)/2$ angles will be "sufficient" to faithfully reconstruct X.

- Given size of the 2D object = $200x200$, the length (in pixel units) of the diagonal of this image = $200 * \sqrt(2)$, and this is twice the radius of the enclosing circle.
- Since our imaging method is circular cone beam projection, we assume a circle to tightly bound the object.
- $R_p = 282.84/2 = 141.4$
- No. of views = $2 * \pi * R_p = 888$ views

Similarly, in the case of 3D, assume the object is enclosed within a sphere of unit radius R. Let $R_v$ denote the radius in voxel units, i.e., number of object-voxels that fit within the radius of the sphere. Hence, in 3D, one way to compute sufficient number of views would be the following:

- Given size of the object scanned = $150x150x100$, the length of the space diagonal of this cuboid = $\sqrt{(100^2 + 100^2 + 150^2)}$
- Since our imaging method is circular cone beam projection, we assume a sphere tightly bounding the object.

- The radius of this sphere in voxel units ($R_v$) will be $\sqrt{(100^2 + 100^2 + 150^2)}/2 = 103$ and, surface area of the sphere $= 4 * \pi * (R_v)^2 = 133320$.
- If we want to pass at least one ray for every surface voxel of this sphere, we will need 133320/2 (since sphere is symmetrical, and ray at 30 deg will pass through same voxels as ray at 30 +180 deg) = 66658 views.

Hence our choice of 900 views for reconstruction of the ground-truth is not exceedingly large. It just happens to be sufficient due to the homogenous structure of the interior of the object (potato).

Since our gold-standard for measuring accuracy of our technique's reconstruction is the ground-truth reconstruction obtained from 900 views (for the potato dataset), we have measured our achieved reduction in views as a fraction of the views used for reconstructing ground-truth.

**R1-C13:** *Hard to notice the features within the ROI in Fig. 15. It may help to present images that zoom into the ROI.*

**Response:** The zoomed-in RoI is now shown in Fig. 16.

**R1-C14:** *In section V (B), how was the cross-validation or empirical selection of $\lambda_1$ and $\lambda_2$ done? Did you use SSIM as a comparison metric? Did you use any particular object prior image for cross-validation? If so, which image did you use?*

**Response:** In our earlier submission, $\lambda_1$ and $\lambda_2$ values were chosen by reconstructing the test volume by sweeping through a range of values and selecting the set of values that gave a reasonable visual quality. In the current version of the paper, we had first chosen $\lambda_1$ as that value which maximizes the SSIM of the l1-ls reconstruction. This value was retained for uniform and spatially-varying prior based reconstruction as well. $\lambda_2$ was chosen clairvoyantly, i.e., based on the value that gave the best visual results on the test.

**R1-C15:** *You are allowed to submit documents or media as supplementary material. Instead of posting them on a dropbox, you could submit them as supplementary material to the journal. You could also use supplementary material to add additional text or figures to reduce number of pages in main paper.*

**Response:** Thank you for the suggestion. We have now submitted the supplementary material on the publication portal.

**R1-C16:** *To save space and if you feel comfortable, you could move the section III (A) (1) on patch based techniques to supplementary since it is not a contribution of your paper.*

**Response:** In order to save space and for greater clarity, we have now moved the discussion on patch-based techniques to supplementary material.

# 4 Response to Reviewer 2

(We have numbered these comments as **R2-Cx**, and italicized for clarity)

**R2-C1.** *I have one main question concerning Table 1, where I notice FDK gives for the ROI much better SSIM value than the variants of the proposed method. What is the reason of this? Why can then the authors state that their method is better than FDK? A bit more detailed explanation would be satisfactory.*

**Response:** We hypothesize that because the volume of this dataset (potato) is very homogenous (without intricate structures), the test measurements capture sufficient information about the new regions even with 5% views and hence FDK performs very well in the exact region of interest (shown as red RoI in Fig. 19(a)). However, as we expand the region of interest (green and cyan RoI in Fig. 19a), the prior-based methods perform better than FDK (please refer Table 8) in that they minimize the sub-sampling artefacts. We have now discussed this in Section 8. B - "Reconstruction of homogenous data". Had the new region been embedded within an intricate background (like the sprouts and okra), its reconstruction does give rise to artefacts as seen in Figs.13-b and 15-b.

Beside this, I only have minor comments:

**R2-C2**. *P1, L49, right panel: "a" should not be typeset in italic.*

**Response:** The italicized 'a' here is used to emphasize the selection of 'any one' particular template amongst the many available ones. So, we retain the typeset on "a" and for greater clarity, italicize the word 'particular' as well . The statement now reads: "there is the key issue of choice of a particular previously scanned object among the many previously acquired scans."

**R2-C3.***P2, L31, left panel: The subtitle should be typeset in bold.*

**Response:** Accepted. The typeset has been changed.

**R2-C4.** *P2, L50, left panel: "reconstruction-" should be corrected to "reconstruction -"*

**Response:** Accepted. This has been corrected.

**R2-C5.***P5, L10, left panel: "Eq 6" should be corrected to "Eq. 6" (same in P6, L54, right panel*

**Response:** Accepted. This has been corrected, and the equation has now been moved to supplementary material (Eq.2).

**R2-C6.** *P6, L17, left panel: "i.e" should be replaced with "i.e." (correct each further occurrence*

**Response:** As per the conventional use, "i.e" has now been replaced by "i.e.," in all occurrences in the paper.

**R2-C7.** *P6, L42, left panel: "Eqn. 8" should be corrected "Eq. 8" (same in P7, L19, left panel*
**Response:** Accepted. These have been corrected.

**R2-C8.** *Caption of Fig. 8: The correct form is (a)-(f) and (g)*
**Response:** Accepted. This has been corrected.

**R2-C9.** *P6, L40, right panel: Subtitle should be typeset in bold*
**Response:** Accepted. The typeset has been changed.

**R2-C10.** *P7, L3, and L7: "proposed methods" ¡–¿ "proposed method". Choose one.*
**Response:** In Sec.6, we use "Proposed method" to refer to uniform prior-based method, and in Sec.8 we use "Proposed method" to refer to spatially varying prior-based method.

**R2-C11.** *Caption of Figs 10, 12, and 13: "Unselective prior" should be typeset in two lines instead of three.*
**Response:** Accepted. This has been typeset.

**R2-C12.** *P7, L50, left panel: Replace "Figs. 9" with "Fig. 9"*
**Response:** Accepted. This has been corrected.

**R2-C13.** *P7, L28, right panel: do not use bold typesetting here*
**Response:** Accepted. This has been corrected.

**R2-C14.** *Caption of Fig. 15: "Letters missing"*
**Response:** Accepted. This has been corrected.

**R2-C15.** *P8, L49, right panel: Correct "shows" to "show"*
**Response:** Accepted. This has been corrected.

**R2-C16.** *Caption of Fig. 16: "slice-8" should be replaced with "image-8"*
**Response:** Accepted. This has been corrected.

**R2-C17.** *P9, L20, right panel: Move [16] to the second place.*
**Response:** Accepted. This had been corrected.

**R2-C18.***P9, L34, right panel: "[38](FDK" space is missing*
**Response:** Accepted. This has been corrected.

**R2-C19.***P9, L36, right panel: "in-dept" should be corrected to "in-depth", "in" should be corrected to "is"*
**Response:** Accepted. These have been corrected.

**R2-C20.** *Caption of Fig. 17: "weight maps" should be corrected to "weights maps"*

**Response:** Accepted. This has been corrected to 'weights-maps' in all occurrences in the paper.

**R2-C21.** *Acknowledgement: "Dr.Andrew Kingston" space is missing*

**Response:** Accepted. This has been corrected.