Multi-Task Learning for Simultaneous Anatomical Segmentation and Lesion Localization in Three-Dimensional Medical Imaging

Enzio Lattuada
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
enziolattuada95@unr.edu

Abstract

Three-dimensional medical imaging supports diagnosis, treatment planning, and disease monitoring, yet anatomical segmentation and lesion localization are often treated as separate computational tasks. This separation creates redundant feature extraction, inconsistent spatial reasoning, and evaluation procedures that obscure whether one task benefits or harms the other. This manuscript develops a technically grounded multi-task learning framework for simultaneous anatomical segmentation and lesion localization in volumetric images. The central contribution is a coupled formulation in which a shared three-dimensional encoder supports a dense anatomical segmentation decoder and a sparse lesion heatmap decoder, while a segmentation-conditioned inference rule reweights lesion evidence by anatomical plausibility without imposing a hard mask. The framework specifies volumetric inputs and outputs, task-specific losses for class-imbalanced segmentation and point-supervised lesion localization, uncertainty-based loss balancing, and an explicit analysis of task conflict through gradient direction. It also defines a separate but coordinated evaluation logic that distinguishes anatomical overlap, boundary accuracy, lesion detection sensitivity, false positives per volume, and localization distance. The analysis identifies a key asymmetry: segmentation errors can attenuate lesion evidence near anatomical boundaries, whereas lesion sparsity and label ambiguity can weaken shared representations if localization gradients are allowed to dominate. These mechanisms imply that coupling is most defensible when it is soft, uncertainty-aware, and evaluated task by task. The formulation provides a reproducible methodological basis for joint volumetric analysis and clarifies the conditions under which shared learning is technically preferable to independent pipelines.

Keywords

multi-task learning; three-dimensional medical imaging; anatomical segmentation; lesion localization; uncertainty weighting; volumetric deep learning

1. Introduction

Volumetric medical imaging produces spatially rich data in which anatomy and pathology are interleaved. Computed tomography, magnetic resonance imaging, and positron emission tomography can reveal organ boundaries, vascular structures, and focal abnormalities within a common coordinate system. Clinical workflows nevertheless tend to decompose analysis into separate operations: delineating organs or tissues, identifying lesions, measuring burden, and reporting locations. This decomposition is practical, but it duplicates computation and can produce inconsistent spatial evidence. A lesion may be detected outside a segmented organ region, or an anatomical segmentation may exclude the very abnormal tissue that should be

localized. Joint modeling offers a way to address this inconsistency, provided the relationship between dense anatomical labels and sparse lesion annotations is formulated carefully.

Encoder-decoder convolutional networks established dense pixel-wise segmentation for biomedical images [1]. Their extension to three-dimensional kernels made volumetric segmentation practical [2], and self-configuring pipelines improved reproducibility across heterogeneous imaging protocols [3]. Multi-task learning extends this architecture by sharing representations among related outputs [4]. A common motivation is statistical efficiency: auxiliary tasks can regularize shared features and reduce overfitting when labels are scarce. Loss balancing can be made adaptive by modeling task uncertainty [5], architectural coupling can be controlled through cross-stitch units [6], and task relationships can be treated as a multi-objective optimization problem [7]. Medical imaging surveys have repeatedly noted that segmentation, detection, and classification differ in label density, noise structure, and clinical cost [8]. Ambiguity in segmentation targets further complicates joint learning because annotators may disagree on boundaries, an issue addressed by probabilistic segmentation models [9].

The specific problem considered here is simultaneous anatomical segmentation and lesion localization in three-dimensional images. The inputs are volumetric scans, the anatomical output is a voxel-wise label map, and the localization output is a spatial map indicating lesion presence or lesion centers. The two tasks are related but not identical. Anatomical segmentation is dense, class-imbalanced, and evaluated with overlap or boundary metrics. Lesion localization is sparse, often point-supervised, sensitive to small structures, and evaluated with detection or distance criteria. A naïve shared model can therefore learn features that serve the dominant task while underfitting the sparse task, or it can let sparse lesion gradients distort anatomical boundaries. The central contribution of this manuscript is not a claim of demonstrated performance but a complete methodological framework that specifies how to couple the tasks, how to balance their objectives, how to reason about error propagation, and how to evaluate joint behavior without conflating anatomical quality with lesion detection.

2. Problem Formulation and Technical Context

Let a volumetric image be defined on a discrete grid and accompanied by an anatomical label map and a set of lesion locations. The basic data object is

where X is the three-dimensional input volume, Ω is the voxel grid, H, W, and D denote spatial dimensions, C is the number of anatomical classes, y assigns one anatomical label to each voxel, and L contains K lesion locations. The locations may be centroids, point annotations, or small seed regions rather than full lesion masks. This distinction matters because point supervision changes the localization problem from dense lesion segmentation to detection-oriented heatmap estimation. The formulation can accommodate full lesion masks when available, but it does not assume them.

Volumetric medical images introduce technical conditions absent from natural-image benchmarks. Voxel spacing is often anisotropic, intensity distributions vary across scanners and protocols, and lesions can occupy only a few voxels in one plane while spanning many slices in another. Class imbalance is severe. Highly unbalanced segmentation has motivated generalized Dice objectives and related overlap losses [10], while volumetric convolutional designs have shown how three-dimensional context can be encoded directly rather than by slice-wise processing [11]. Evaluation is equally consequential: overlap metrics, boundary

distances, and volume-based measures capture different failure modes [12]. Public benchmarks such as liver tumor segmentation and brain tumor segmentation illustrate the heterogeneity of lesion tasks and annotation conventions [13,14]. These conditions imply that anatomical segmentation and lesion localization should not be optimized with a single undifferentiated loss or assessed with a single aggregate score.

The anatomical and lesion tasks also differ in their admissible spatial support. Anatomical labels describe expected tissue classes, whereas lesions can be focal, multifocal, infiltrative, or extra-anatomical. A model that treats anatomical segmentation as a hard prior for lesion localization may suppress true positives outside an under-segmented organ. Conversely, a model that ignores anatomy entirely may produce lesion heatmaps concentrated on normal structures with intensity patterns similar to pathology. The framework therefore treats anatomical context as a soft conditioning signal and retains an explicit localization pathway that can override anatomical expectations when evidence is strong.

3. Methodological Framework

3.1 Shared volumetric representation

The proposed architecture uses a shared three-dimensional encoder and two task-specific decoders. The encoder maps an input volume to a multi-scale latent representation. The segmentation decoder produces voxel-wise anatomical logits, while the localization decoder produces a normalized lesion heatmap. The forward mapping can be written as

$$z = E_\theta(X), \hat{y} = softmax(S_\varphi(z)), \hat{h} = H_\psi(z)$$

where E_θ is the shared encoder with parameters θ, S_φ is the anatomical segmentation decoder with parameters φ, H_ψ is the localization decoder with parameters ψ, z denotes the encoded feature map, ŷ denotes voxel-wise anatomical class probabilities, and ĥ denotes the lesion heatmap. The heatmap is defined on the same grid as the input or on a fixed downsampled grid with coordinates mapped back to image space. A shared encoder is appropriate because low- and mid-level features such as edges, intensity transitions, and tissue texture are relevant to both tasks. Task-specific decoders preserve output semantics and prevent the dense anatomical label space from directly dictating the localization output.

Because full volumetric processing can exceed memory limits, training is formulated on overlapping patches. Patch sampling should not be uniform. It should oversample volumes containing lesions while retaining negative patches that contain anatomically challenging regions. The encoder can use anisotropic kernels or spacing-aware normalization so that physical distances remain meaningful across acquisitions. Skip connections preserve fine spatial detail for anatomical boundaries and small lesions, but the localization decoder should not simply reuse segmentation logits as its input; doing so would make localization a deterministic function of segmentation and eliminate complementary evidence from the shared features.

3.2 Anatomical segmentation branch

The segmentation branch predicts a class probability for every voxel. Its objective combines cross-entropy with an overlap term to address class imbalance. For a generic anatomical class, with $y_v \in \{0,1\}$ denoting the reference label at voxel v and $\hat{y}_v$ denoting the predicted probability, the segmentation loss is

$$L_seg = -\sum_v y_v \log \hat{y}_v + \lambda(1 - 2\sum_v y_v \hat{y}_v / (\sum_v y_v + \sum_v \hat{y}_v))$$

where v ranges over the sampled voxel set, the first term penalizes voxel-wise classification error, the second term penalizes low overlap, and λ controls the contribution of the overlap term. In multi-class settings, the expression is evaluated per class and averaged, or the generalized Dice formulation can be used to weight classes inversely to their volume [10]. The overlap term is especially important for small anatomical structures, although it can become unstable when a class is absent from a patch. For that reason, class presence should be tracked during patch construction, and empty-class terms should be handled consistently rather than silently omitted.

The segmentation branch serves two purposes in the joint model. First, it produces the anatomical output required by the clinical task. Second, it supplies a spatial prior for localization conditioning. This second role requires segmentation probabilities rather than hard labels. A hard anatomical mask would turn any segmentation error into an irreversible localization exclusion, whereas a probability map allows the localization branch to retain evidence in uncertain regions. The segmentation decoder should therefore be trained with adequate capacity and evaluated independently before its output is used as a conditioning signal.

3.3 Lesion localization branch

The localization branch predicts a heatmap $\hat{h}$ over the volume. Point-supervised lesions are converted into spatial target maps by placing a Gaussian kernel around each lesion coordinate. For lesion k at coordinate p_k, the target contribution is

$$g_k(v) = exp(-\|v - p_k\|^2/(2\sigma_k^2))$$

where v is a voxel coordinate, p_k is the annotated lesion location, and σ_k controls the spatial spread of the target. When multiple lesions are present, the target map is g(v) = max_k g_k(v), which prevents overlapping Gaussians from creating artificially large activation values. The spread σ_k should reflect annotation uncertainty and physical lesion scale rather than a single global constant. A very small σ produces an almost one-hot target and makes optimization brittle; a very large σ blurs adjacent lesions and degrades localization specificity.

The localization loss is a weighted binary cross-entropy over the heatmap:

$$L_loc = -\sum_v w(v)\left[g(v)\log\hat{h}(v) + \left(1 - g(v)\right)\log\left(1 - \hat{h}(v)\right)\right]$$

where g(v) is the target heatmap and w(v) = 1 + κg(v) is a voxel weight with κ ≥ 0. The weight increases the contribution of lesion-bearing voxels and mitigates the extreme foreground-background imbalance of point-supervised targets. The loss should be interpreted as a detection surrogate, not as a calibrated lesion probability, because the Gaussian target encodes spatial tolerance rather than the true extent of disease. If full lesion masks are available, the localization branch can be modified to predict lesion masks and the Gaussian target can be replaced by an overlap loss. The point-supervised formulation is retained here because lesion masks are often unavailable in volumetric archives.

3.4 Coupling segmentation and localization

The two branches interact through shared features and through a soft inference coupling. Let R denote the set of anatomical classes that define a plausible region for the lesion task. For example, R may contain liver, kidney, or brain tissue depending on the application. The conditioned localization map is

$$\tilde{h}(v) = \hat{h}(v)(1 + \alpha \sum_{c \in R} \hat{y}_c(v))$$

where ŷ_c(v) is the probability that voxel v belongs to anatomical class c, and α ≥ 0 controls the strength of anatomical conditioning. When α = 0, localization is independent of anatomy. When α is large, the heatmap is strongly amplified inside the segmented region. The multiplicative form is deliberately soft: it raises the score of lesions within anatomically plausible tissue without setting scores outside the region to zero. This design reflects the asymmetry between false positives and false negatives. Suppressing a true lesion because of an under-segmented organ boundary is more damaging than retaining a modest score in an anatomically unexpected location.

The coupling also has a training-time dimension. Gradients from both losses update the shared encoder, so features useful for anatomical boundaries can influence localization and vice versa. This interaction can regularize the encoder when lesion labels are sparse. It can also create negative transfer if the segmentation task dominates and the encoder allocates capacity to normal anatomy that is irrelevant to lesion appearance. The uncertainty-weighted objective described next is intended to reduce, not eliminate, this risk.

3.5 Multi-task objective and uncertainty

The joint objective combines the segmentation and localization losses with learned task uncertainty terms:

$$L = \sum_{t \in \{seg, loc\}} (1/(2s_t^2))L_t + \log s_t + \lambda_r \|\Theta\|^2$$

where t indexes the task, L_t is the corresponding task loss, s_t > 0 is a learned uncertainty parameter for task t, Θ denotes all trainable parameters, λ_r controls weight decay, and the logarithmic term prevents unbounded growth of s_t. This formulation follows the homoscedastic uncertainty weighting principle in which a task with higher estimated uncertainty receives a smaller effective loss weight [5]. It is preferable to fixed weights when the scale of the lesion loss varies across volumes because the number of lesions and the size of the sampled patch change. The parameter s_t should be constrained or parameterized in log space to avoid numerical instability.

Uncertainty weighting does not resolve task conflict by itself. It adjusts the relative magnitude of gradients but does not guarantee that their directions are compatible. For this reason, the localization decoder should include dropout-based or ensemble-based uncertainty estimation at inference [15,16]. Predictive entropy can flag voxels where the heatmap is unstable, and the anatomical segmentation can be similarly summarized. These uncertainty maps are not substitutes for external validation; they provide an inspection mechanism for identifying cases in which the joint model is confident for the wrong reason.

3.6 Optimization and inference

Training can proceed with patch-based stochastic optimization, balanced lesion sampling, and augmentation in physical space. Intensity augmentation should preserve tissue contrast relationships, while geometric augmentation should respect voxel spacing and avoid transformations that create anatomically implausible configurations. The segmentation and localization losses should be monitored separately during training. Early stopping based only on the joint loss can conceal negative transfer, because a decrease in the dominant segmentation loss may offset a degradation in localization.

At inference, the anatomical segmentation is computed first, followed by the localization heatmap and the soft conditioning in Equation (6). Lesion candidates are obtained by local

maxima detection above a threshold τ, with non-maximum suppression in three dimensions. The threshold should be selected on validation data according to the clinical objective, not fixed by convention. If the application prioritizes high sensitivity, τ is lowered and false positives per volume are reported explicitly. If the application prioritizes precision, τ is increased and the resulting sensitivity loss is quantified. The soft conditioning parameter α can also be selected on validation data, but it should not be tuned on the test set.

4. Evaluation Logic and Error Analysis

4.1 Metrics and validation

Joint evaluation must report anatomical and localization performance separately before reporting any combined score. Anatomical segmentation should be assessed with overlap metrics and boundary distances [12]. Dice and Jaccard coefficients summarize regional agreement, while the 95th percentile Hausdorff distance reflects boundary error and is less forgiving of small but clinically relevant misplacements. Because these metrics behave differently on small structures, results should be stratified by anatomical class and lesion size. Lesion localization should be assessed with detection sensitivity, false positives per volume, and localization distance. A detection is counted when a predicted maximum falls within a distance δ of a reference lesion and exceeds threshold τ; distance error can be summarized as the median and upper quantile over matched lesions. These criteria are more informative than voxel-wise overlap for point-supervised targets.

Validation should be patient-level rather than slice-level or patch-level. Patch-level splitting leaks adjacent anatomy and lesion context across training and test sets, inflating apparent performance. External validation on a different scanner, institution, or acquisition protocol is necessary before any claim of transportability. The evaluation should also include calibration of the heatmap and segmentation probabilities, because a model with adequate discrimination can still be poorly calibrated. Calibration matters when thresholds are used to decide whether a lesion is reported or an anatomical boundary is accepted.

4.2 Error modes

The joint formulation creates specific failure modes. The first is anatomical gating error. If the segmentation branch under-segments the organ containing a lesion, the factor $\sum_{c \in R} \hat{y}_c(v)$ becomes small near the lesion and the conditioned heatmap $\tilde{h}(v)$ is attenuated. This error is not symmetric: over-segmentation usually enlarges the plausible region and retains more lesion evidence, whereas under-segmentation can suppress true positives. The soft form of Equation (6) reduces but does not remove this risk. The second is lesion-induced anatomical distortion. A large lesion may displace or replace normal tissue, causing the anatomical segmentation to assign low probability to the organ class at the lesion location. In that case, anatomical conditioning conflicts with pathology evidence. The localization branch must therefore retain sufficient independent capacity to detect lesions outside the dominant anatomical prior.

The third error mode is sparse-label overfitting. When only a few lesion coordinates are available, the localization decoder can learn to respond to acquisition artifacts, contrast patterns, or annotated-site prevalence rather than lesion biology. Shared representations from segmentation can help, but they can also encourage the localization branch to rely on anatomical context at the expense of lesion-specific texture. The fourth is heatmap ambiguity. A Gaussian target defines a region of tolerance, but it does not specify whether the model

should localize the lesion center, its boundary, or any point within the lesion. This ambiguity affects distance-based evaluation and should be made explicit in annotation guidelines.

4.3 Sensitivity and task conflict

The interaction between tasks can be examined through gradient similarity. For shared parameters θ, the cosine similarity between task gradients is

$$\rho = \langle \nabla_\theta L_{seg}, \nabla_\theta L_{loc} \rangle / (\|\nabla_\theta L_{seg}\| \|\nabla_\theta L_{loc}\|)$$

where ∇_θ denotes the gradient with respect to shared parameters and the inner product measures directional agreement. Positive values indicate that the two tasks push shared features in similar directions; negative values indicate conflict. This quantity is diagnostic rather than decisive, because a negative cosine in one layer does not imply that joint training is harmful overall. It does, however, identify layers in which task-specific adaptation may be preferable to unrestricted sharing. A practical design implication is that early encoder stages can remain shared, while later stages or normalization statistics can be partially task-specific when conflict is persistent.

The conditioning strength α has a clear sensitivity trade-off. Increasing α improves precision when the anatomical segmentation is accurate and the lesion lies within the expected region, but it also raises the cost of segmentation errors. Decreasing α preserves localization independence but weakens the benefit of anatomical context. The Gaussian width σ_k similarly controls the bias-variance trade-off of point supervision. Small values increase localization specificity but make optimization sensitive to coordinate noise; large values improve recall tolerance but merge nearby lesions and reduce distance accuracy. Finally, the uncertainty parameters s_t affect the balance between tasks but should not be interpreted as clinical confidence without calibration.

A related strategy for sparse spatial targets appears in segmentation-guided semi-supervised landmark detection on intraoral three-dimensional scans [17]. That work is relevant because it shows how segmentation cues can guide a localization-like task when annotations are limited, but its target is a set of anatomically defined landmarks rather than a variable lesion heatmap. Lesions can occur outside expected anatomical regions and can vary in size and multiplicity, so segmentation guidance cannot be treated as a hard anatomical constraint. The present formulation therefore uses segmentation only as a soft reweighting factor and retains a separate localization loss and evaluation pathway.

5. Conclusion

This manuscript has developed a multi-task learning framework for simultaneous anatomical segmentation and lesion localization in three-dimensional medical imaging. The framework combines a shared volumetric encoder with task-specific segmentation and localization decoders, a class-imbalanced segmentation objective, a point-supervised Gaussian heatmap objective, soft segmentation-conditioned localization, and uncertainty-based loss weighting. Its central technical argument is that anatomical context is useful for lesion localization only when it is probabilistic and reversible. Hard anatomical masks propagate segmentation errors directly into detection, whereas a soft multiplicative coupling preserves lesion evidence outside uncertain anatomical regions.

The analysis also clarifies why joint evaluation must remain task-specific. Anatomical overlap and boundary metrics measure dense structure delineation, while lesion detection sensitivity, false positives per volume, and localization distance measure sparse pathological

findings. A single aggregate score can conceal negative transfer. Gradient-direction analysis provides a further diagnostic for deciding where sharing is beneficial and where task-specific capacity is warranted. These design choices do not guarantee clinical utility, but they define a coherent and testable methodological basis for joint volumetric analysis. The framework is most appropriate when lesion annotations are sparse, anatomical segmentation is available or feasible, and the clinical objective requires both structural delineation and spatial lesion reporting.

References

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

2. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016 (pp. 424–432). Springer. https://doi.org/10.1007/978-3-319-46723-8_49

3. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

4. Caruana, R. (1997). Multitask learning. Machine Learning, 28(1), 41–75. https://doi.org/10.1023/A:1007379606734

5. Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7482–7491). IEEE. https://doi.org/10.1109/CVPR.2018.00781

6. Misra, I., Shrivastava, A., Gupta, A., & Hebert, M. (2016). Cross-stitch networks for multi-task learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3994–4003). IEEE. https://doi.org/10.1109/CVPR.2016.433

7. Sener, O., & Koltun, V. (2018). Multi-task learning as multi-objective optimization. In Advances in Neural Information Processing Systems (Vol. 31, pp. 527–538). Curran Associates.

8. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

9. Kohl, S. A. A., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J. R., Maier-Hein, K. H., Eslami, S. M. A., Rezende, D. J., & Ronneberger, O. (2018). A probabilistic U-Net for segmentation of ambiguous images. In Advances in Neural Information Processing Systems (Vol. 31, pp. 6965–6975). Curran Associates.

10. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical

Decision Support (pp. 240–248). Springer. https://doi.org/10.1007/978-3-319-67558-9_28

11. Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the Fourth International Conference on 3D Vision (pp. 565–571). IEEE. https://doi.org/10.1109/3DV.2016.79

12. Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Medical Imaging, 15, 29. https://doi.org/10.1186/s12880-015-0068-x

13. Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G. E., Chartrand, G., Lohofer, F., Holch, J. W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdzal, M., Amitai, M. M., Sosna, J., ... & Menze, B. (2023). The liver tumor segmentation benchmark (LiTS). Medical Image Analysis, 84, 102680. https://doi.org/10.1016/j.media.2022.102680

14. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., ... & Van Leemput, K. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging, 34(10), 1993–2024. https://doi.org/10.1109/TMI.2014.2377694

15. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the 33rd International Conference on Machine Learning (PMLR Vol. 48, pp. 1050–1059). PMLR.

16. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In Advances in Neural Information Processing Systems (Vol. 30, pp. 6402–6413). Curran Associates.

17. Liu, S., Jin, J., Li, W., & Yan, X. (2026). Segmentation-Guided Semi-Supervised Landmark Detection on Intraoral 3D Scans. Preprints. https://doi.org/10.20944/preprints202609.1722.v1