Anatomy-Aware Deep Learning for Precise Localization of Liver Tumors in Contrast-Enhanced Computed Tomography

Mauri Lipponen
School of Computing, Clemson University, Clemson, SC, USA.
maurilipponen98@clemson.edu

Abstract

Precise localization of liver tumors in contrast-enhanced computed tomography is a clinically consequential problem that differs from voxel-wise tumor segmentation: a clinically useful system must identify the correct hepatic lesion, estimate its three-dimensional position relative to liver anatomy, distinguish tumor from adjacent vessels and parenchyma, and communicate when a localization is unreliable. Deep segmentation networks have improved liver and tumor delineation, but segmentation output alone does not resolve lesion identity, extrahepatic false positives, phase-dependent enhancement, or respiratory deformation across contrast phases. This manuscript develops an anatomy-aware deep learning framework for tumor localization that couples multiphase computed tomography representation, liver-constrained heatmap prediction, deformable inter-phase registration, centroid estimation, and uncertainty-gated decision making. The central contribution is a technically explicit localization formulation in which dense anatomical priors are not appended as post-processing but are embedded in the model's spatial scoring and training objectives. A distance-transform representation of the liver domain suppresses implausible extrahepatic responses while permitting a controlled boundary tolerance for subcapsular and exophytic lesions. A registration-consistency term links arterial, portal-venous, and delayed-phase predictions, and Monte Carlo dropout supplies an entropy-based reliability signal. The manuscript specifies the component functions, loss terms, assumptions, and boundary conditions, then analyzes likely failure modes including low-contrast lesions, mistimed contrast, adjacent-vessel confusion, small-lesion centroid instability, and inter-phase registration error. The resulting framework provides a defensible methodological basis for precise, uncertainty-aware liver tumor localization and identifies the evaluation endpoints needed before clinical translation.

Keywords

liver tumor localization; contrast-enhanced computed tomography; anatomy-aware deep learning; deformable registration; uncertainty estimation; segmentation-guided detection

1. Introduction

Liver cancer remains a major source of cancer mortality worldwide, and contrast-enhanced computed tomography is central to detection, staging, treatment selection, and image-guided intervention [1]. Hepatocellular carcinoma in particular often arises in a cirrhotic liver with distorted parenchymal architecture, heterogeneous enhancement, and multifocal disease, making lesion identification and spatial characterization more demanding than detection in homogeneous organs [2]. Accurate localization is required for percutaneous ablation, surgical planning, stereotactic radiotherapy, and longitudinal response assessment.

The task is not merely to outline a lesion in one image volume. It requires placing the lesion in a consistent anatomical coordinate system, separating it from vessels and nearby parenchymal abnormalities, and determining whether the predicted position is reliable enough to support a clinical decision.

Deep learning has substantially advanced medical image segmentation. The U-Net architecture established an encoder-decoder design with skip connections that remains influential for biomedical image segmentation [3], and self-configuring pipelines such as nnU-Net have demonstrated the value of systematic preprocessing, network configuration, and post-processing for volumetric tasks [4]. The Liver Tumor Segmentation Benchmark has further standardized large-scale evaluation of liver and tumor delineation and has exposed persistent difficulties in small, low-contrast, and boundary-ambiguous lesions [5]. These developments provide a strong foundation for tumor detection, yet voxel overlap is not equivalent to precise localization. A segmented mask may overlap a lesion while its estimated center is displaced toward a vessel, liver capsule, or adjacent lesion. Conversely, a small but clinically important lesion may be missed even when global segmentation metrics appear acceptable. Multiphase contrast enhancement adds diagnostic information but also introduces deformation and timing variability that must be reconciled through registration and consistency modeling [6].

A second line of work indicates that dense anatomical constraints can support sparse landmark or target localization under limited supervision. Segmentation-guided semi-supervised landmark detection has been explored in intraoral three-dimensional scans as a strategy for linking dense anatomical delineation to sparse positional targets [7]. That work is not a liver study and does not establish hepatic performance, but it supports a general design principle: when sparse localization labels are expensive, dense anatomical structure can regularize where a target is expected to lie. Broader reviews of deep learning in medical image analysis have documented both the promise and the fragility of such models under domain shift, class imbalance, and limited annotation [8]. The gap addressed here is the absence of an explicit, anatomy-aware formulation that couples liver-domain constraints, multiphase enhancement, registration consistency, and uncertainty gating for precise hepatic tumor localization rather than for segmentation alone.

The proposed contribution is a methodological framework in which localization is represented as an anatomy-constrained spatial decision. The framework uses a segmentation backbone to derive liver, vessel, and tumor-related features; converts the liver domain into a distance field that modulates tumor heatmap responses; estimates lesion centroids by spatial expectation; enforces cross-phase consistency after deformable registration; and applies an entropy-based rule that can withhold or flag uncertain localizations. The manuscript develops the formulation, analyzes its assumptions and failure modes, and specifies how the framework should be assessed without presenting fabricated empirical results. Its scientific significance lies in making anatomical reasoning, inter-phase geometry, and uncertainty explicit parts of the localization model rather than incidental post-processing steps.

2. Imaging and Localization Problem

2.1 Contrast phases and attenuation model

Contrast-enhanced computed tomography does not provide a single stable observation of a liver lesion. Arterial-phase images often emphasize hypervascular tumor components, portal-venous images improve parenchymal and venous contrast, and delayed images can

reveal washout or capsular features. Respiratory motion, patient positioning, and contrast timing cause spatial and intensity differences across phases. An idealized phase-specific observation at voxel x can be written as

$$I_p(x) = \mu_t(x) + \beta_p \kappa(x) + n_p(x)$$

where p denotes the contrast phase, I_p(x) is the measured attenuation in Hounsfield units, μ_t(x) is the baseline attenuation associated with tissue class t, κ(x) is a phase-dependent contrast uptake component, β_p is the phase weighting, and n_p(x) represents noise and acquisition artifacts. The expression assumes approximate spatial alignment; before registration, x in one phase does not necessarily correspond to the same anatomical point in another phase. The model is deliberately simple. Its purpose is to make explicit why a lesion may be conspicuous in one phase and nearly isointense in another, and why a localization model should not treat multiphase input as interchangeable channels without accounting for deformation.

2.2 Localization as an anatomical decision

Let Ω denote the liver domain in a reference volume, and let c_k be the three-dimensional center of the kth tumor. A localization prediction should satisfy three conditions: it should lie within a plausible hepatic region, it should be stable across contrast phases after registration, and it should have low predictive uncertainty. These conditions are not fully captured by Dice overlap or by voxel-wise sensitivity. A useful localization output is therefore a coordinate, a confidence value, and an anatomical context descriptor indicating proximity to vessels, capsule, and segment boundaries. The framework treats the liver domain as a soft constraint rather than an absolute mask. Subcapsular and exophytic lesions can extend beyond a conventional liver mask, so the model must allow a boundary tolerance while still penalizing distant extrahepatic responses. This distinction is central to avoiding both missed peripheral lesions and false localizations in the stomach, bowel, or adjacent soft tissue.

3. Methodological Framework

3.1 Anatomy-aware segmentation backbone

The framework begins with a volumetric segmentation backbone that receives multiphase patches and produces liver, vessel, and lesion-related feature maps. The backbone can follow a U-Net-style encoder-decoder design [3] with multiscale skip pathways such as UNet++ [9] or a volumetric V-Net formulation [10]. The specific architecture is less important than the requirement that the network preserve fine spatial detail while aggregating sufficient context to distinguish lesions from vessels and cirrhosis-related parenchymal heterogeneity. The liver mask defines Ω. Vessel masks are used to compute a vessel proximity field, because lesions adjacent to portal or hepatic veins are a common source of centroid displacement and false-positive detection. The liver distance field D_Ω(x) is zero inside Ω and increases with Euclidean distance outside Ω. A separate signed distance to the vessel tree can be included as an auxiliary channel, but it is not used as a hard exclusion rule because tumor-vessel contact is clinically common.

3.2 Segmentation-guided tumor heatmap and centroid estimation

The localization head predicts an anatomy-aware heatmap h_θ(x|V) over the image volume V. A practical formulation is

$$h_\theta(x|V) = \sigma\big(g_\theta(x|V)\big)\exp\big(-\lambda D_\Omega(x)\big)$$

where g_θ(x|V) is the raw network logit for tumor presence, σ(·) is the logistic function, θ denotes model parameters, and λ ≥ 0 controls the strength of the extrahepatic penalty. Inside the liver, D_Ω(x) is zero and the heatmap is determined primarily by image and segmentation features. Outside the liver, the exponential term attenuates predictions smoothly rather than imposing a brittle binary mask. The parameter λ should be chosen so that a small tolerated boundary zone remains active while distant extrahepatic responses are strongly suppressed. This formulation directly connects dense liver anatomy to the sparse localization target and is conceptually related to segmentation-guided landmark reasoning in other anatomical domains [7], while adapting the idea to phase-varying hepatic computed tomography.

The tumor centroid is estimated by spatial expectation:

$$\hat{c_k} = \sum_x x h_\theta \frac{(x|V)}{\sum_x h_\theta(x|V)}$$

where Ĉ_k is the predicted center of lesion k and the sum is taken over voxels with non-negligible heatmap mass. For a solitary lesion, this expectation provides a stable coordinate that integrates evidence from the entire predicted heatmap. For multifocal disease, however, a single expectation can collapse separate lesions into an intermediate position. The model should therefore include a lesion-separation mechanism, such as peak detection on the heatmap followed by connected-component assignment or a learned instance head. The centroid is not intended to replace the segmentation mask; it provides a compact localization variable that can be checked against anatomy, phase consistency, and uncertainty.

Supervised training would combine segmentation and localization objectives. A representative localization loss is

$$L_loc \; = \; \sum_\{k = 1\}\^\{K\} \, w_k \, ||c_k \; - \; \hat{c}_k||_2\^2$$

where c_k is the reference lesion center, w_k weights lesions according to annotation confidence or clinical importance, and K is the number of labeled lesions in the patch. The total objective would also include a Dice or cross-entropy segmentation term and regularization terms. The weighting is important because small lesions may have less stable manual centroids, while large lesions may dominate voxel-wise losses. Liver tumor segmentation studies using fully convolutional networks and object-based post-processing have shown that candidate generation and false-positive suppression are critical components of hepatic lesion analysis [11]. Cascaded or two-stage designs can improve liver and tumor delineation by separating organ localization from lesion segmentation [12]. The present framework adopts that separation but retains an explicit coordinate-level objective so that localization accuracy is optimized directly rather than inferred only from mask overlap.

3.3 Cross-phase registration and consistency

Multiphase input requires spatial correspondence. Deformable registration seeks a transformation T that aligns a moving phase to a reference phase while preserving anatomical plausibility. A standard regularized registration energy is

$$E(T) \; = \; \sum_x \, ||I_a(x) \; - \; I_b(T(x))||_2\^2 \; + \; \rho \, ||\nabla(T(x) - x)||_2\^2$$

where I_a and I_b are two phase volumes, T(x) is the deformation mapping, and ρ controls smoothness. The first term encourages intensity agreement and the second penalizes irregular deformation. Registration of liver computed tomography is complicated by contrast enhancement, which violates the constant-intensity assumption, and by respiratory motion,

which can produce sliding interfaces near the diaphragm and abdominal wall. Registration should therefore use a liver-focused mask, a multi-resolution schedule, and a deformation model that permits realistic hepatic motion without unrealistic folding. The principles of non-rigid registration are well established [6], but their integration with lesion localization requires an additional consistency constraint.

If ĉ_k^p is the predicted centroid in phase p and T_{p→q} maps phase p coordinates to phase q, cross-phase consistency can be expressed as

$$L_cons = \sum_\{k, p \neq q\} ||T_\{p \to q\}(\hat{c}_k\hat{}p) - \hat{c}_k\hat{}q||_2\hat{}2$$

where the sum runs over lesions and phase pairs. This term penalizes predictions that disagree after registration. It should be weighted by registration confidence, because a large residual may indicate either a localization error or a registration failure. A robust implementation would estimate registration uncertainty from local intensity mismatch and deformation regularity, then down-weight consistency penalties in regions where alignment is unreliable. The consistency term is not intended to force identical predictions across phases when true enhancement differences alter visibility. Its role is to stabilize the anatomical coordinate while allowing phase-specific evidence to contribute to detection.

3.4 Uncertainty and decision gating

A localization system that cannot signal ambiguity is poorly suited to clinical use. Monte Carlo dropout provides a practical approximation to Bayesian model uncertainty by applying dropout at inference and collecting multiple stochastic predictions [15]. For a heatmap model, the mean probability at a candidate centroid can be written as $\bar{h}(\hat{c})$, and predictive entropy is

$$H(\hat{c}) = -\bar{h}(\hat{c}) \log \bar{h}(\hat{c}) - (1 - \bar{h}(\hat{c})) \log(1 - \bar{h}(\hat{c}))$$

where H(ĉ) is the binary entropy of the mean prediction. High entropy indicates that repeated stochastic passes disagree about tumor presence at the predicted location. Entropy alone does not distinguish aleatoric ambiguity caused by poor image contrast from epistemic uncertainty caused by unfamiliar anatomy or acquisition characteristics, a distinction emphasized in Bayesian deep learning for computer vision [16]. The framework should therefore combine entropy with anatomical plausibility and registration consistency. A simple acceptance rule is

$$A = 1 \; if \; H(\hat{c}) \leq \tau_H \; and \; D_\Omega(\hat{c}) = 0 \; else \; 0$$

where A is a binary acceptance indicator and τ_H is a threshold on entropy. In practice, the hard condition D_Ω(ĉ) = 0 should be replaced by a tolerated distance interval at the liver boundary, and the threshold τ_H should be calibrated on held-out data rather than fixed by convention. The output of the gate is not a diagnosis. It is a triage signal indicating whether the localization is sufficiently stable for automated reporting, requires manual review, or should be discarded.

4. Analytical Assessment and Failure Modes

The framework's behavior can be analyzed through the error modes it is designed to address. Low-contrast lesions in fatty or cirrhotic liver are a primary challenge. When the lesion-to-parenchyma difference is small, the heatmap may be diffuse, and the centroid expectation may drift toward adjacent vessels or fibrosis. The anatomy penalty does not solve this problem because the lesion is intrahepatic. Cross-phase consistency can help if the lesion is visible in at least one phase, but it can also propagate a confident error from one phase to

another if registration is poor. A practical safeguard is to require evidence from at least two phases or a high-confidence single-phase prediction before accepting a localization.

Subcapsular and exophytic lesions challenge hard anatomical constraints. If the liver mask is too tight, a peripheral lesion may be attenuated by the distance penalty and localized inward. The soft penalty in the heatmap formulation reduces this risk because the exponential term remains nonzero outside the liver, but λ must be tuned with boundary lesions in mind. A useful design is an asymmetric distance field: a narrow tolerated band outside the liver receives a modest penalty, while more distant extrahepatic regions receive a strong penalty. Vessel-adjacent lesions present a different failure mode. The vessel proximity channel may improve discrimination when a lesion is embedded in or abuts a vessel, but it can also encourage the model to mistake a vessel cross-section for a tumor if enhancement and shape cues are ambiguous.

Registration error is a structural limitation of multiphase consistency. Contrast enhancement changes intensities, and respiratory deformation can be nonuniform across the liver. A residual after registration may reflect true anatomical motion, contrast timing differences, or deformation error. Weighting the consistency term by local registration confidence is therefore necessary. The framework should also avoid using consistency as a substitute for detection evidence. A lesion that is invisible in the portal-venous phase should not be forced into that phase if the registration and phase-specific heatmap do not support it.

Small-lesion centroid instability is another important boundary condition. For a lesion only a few voxels wide, the spatial expectation can change substantially with small changes in heatmap mass. A peak-based estimator or a local weighted centroid within a bounded window may be more stable than a global expectation. Multifocal disease likewise requires instance-aware outputs. Without lesion separation, the centroid of two nearby lesions can fall between them, producing a localization that is anatomically plausible but clinically wrong. Evaluation should therefore include lesion-level hit rates at clinically meaningful distance thresholds, extrahepatic false-positive rates, multi-lesion matching, and calibration of the uncertainty gate. Dice and voxel-wise sensitivity remain useful but insufficient endpoints.

5. Technical Discussion and Implementation Considerations

The framework's data requirements are substantial. It needs multiphase computed tomography with consistent contrast timing, liver and vessel annotations, lesion-level centroids or bounding boxes, and enough boundary and vessel-adjacent cases to expose the failure modes described above. Annotation protocols should define whether centroids are geometric centers of the visible lesion, centers of the enhancing component, or clinically selected ablation targets, because these definitions are not equivalent. If the clinical task is ablation planning, the target may be a safe zone rather than the lesion center. The localization objective must match the intended use.

Transfer learning and fine-tuning can reduce the annotation burden, but medical imaging models are sensitive to acquisition and population shift. Prior work has shown that fine-tuning pretrained networks can outperform or match full training in some medical imaging settings, while performance depends strongly on domain similarity and data scale [17]. A liver localization system should therefore be validated across scanners, contrast protocols, slice thicknesses, and patient populations. Reviews of deep learning segmentation emphasize that preprocessing, class imbalance, and evaluation design are not peripheral details but determinants of clinical usefulness [18]. The same is true for localization: a model that reports

a coordinate without a reliability estimate or a lesion-level error analysis is not a complete clinical tool.

Implementation should also consider computational cost. Volumetric multiphase input, multiscale segmentation, deformable registration, and Monte Carlo inference all increase memory and latency. A staged design can mitigate this burden: run a fast liver and lesion candidate detector, apply registration and consistency checks only to candidate regions, and reserve stochastic uncertainty estimation for ambiguous cases. This cascade preserves the anatomy-aware logic while avoiding unnecessary full-volume computation. The final output should include the coordinate, lesion-level confidence, phase-wise evidence, distance to liver boundary and vessels, and the acceptance state. Such an output supports human review and makes the model's reasoning auditable without pretending that a coordinate alone is a clinical conclusion.

The framework also has boundaries that should be stated plainly. It is not a diagnostic classifier and does not replace contrast-phase interpretation by a radiologist. It assumes that the liver can be segmented with sufficient accuracy to define Ω and that registration is meaningful for the phases under consideration. It does not resolve motion artifacts that corrupt all phases, and it cannot recover lesions that are entirely invisible on the available images. These are not peripheral limitations; they define the conditions under which anatomy-aware localization is valid.

6. Conclusion

This manuscript has developed an anatomy-aware deep learning framework for precise localization of liver tumors in contrast-enhanced computed tomography. The framework treats localization as an anatomical and geometric decision rather than a by-product of voxel segmentation. It embeds a liver-domain distance penalty in a tumor heatmap, estimates lesion centroids from spatial evidence, couples multiphase predictions through deformable registration and consistency loss, and gates outputs with entropy-based uncertainty and anatomical plausibility. The analysis identifies the principal failure modes that determine clinical usefulness: low-contrast lesions, subcapsular and exophytic boundaries, vessel adjacency, registration error, small-lesion centroid instability, and multifocal disease. It also specifies the evaluation endpoints that matter for localization, including lesion-level hit rates, extrahepatic false positives, multi-lesion matching, and uncertainty calibration. The contribution is not a claim of validated clinical performance. It is a technically explicit and internally consistent basis for building and assessing anatomy-aware localization systems, with the central principle that dense hepatic structure, inter-phase geometry, and reliability estimation should be part of the model's localization logic rather than afterthoughts.

References

1. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660

2. Villanueva, A. (2019). Hepatocellular carcinoma. New England Journal of Medicine, 380(15), 1450–1462. https://doi.org/10.1056/NEJMra1713263

3. 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

4. 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

5. Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., ... Menze, B. (2023). The Liver Tumor Segmentation Benchmark (LiTS). Medical Image Analysis, 84, 102680. https://doi.org/10.1016/j.media.2022.102680

6. Crum, W. R., Hartkens, T., & Hill, D. L. G. (2004). Non-rigid image registration: Theory and practice. The British Journal of Radiology, 77(suppl_2), S140–S153. https://doi.org/10.1259/bjr/25329214

7. 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

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. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2020). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging, 39(6), 1856–1867. https://doi.org/10.1109/TMI.2019.2959609

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

11. Chlebus, G., Schenk, A., Moltz, J. H., van Ginneken, B., Hahn, H. K., & Meine, H. (2018). Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. Scientific Reports, 8, 15497. https://doi.org/10.1038/s41598-018-33860-7

12. Christ, P. F., Ettlinger, F., Grun, F., Lipkova, J., Kaissis, G., Ziegelmayer, S., ... Menze, B. H. (2017). Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv. https://arxiv.org/abs/1702.05970

13. Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis, 36, 61–78. https://doi.org/10.1016/j.media.2016.10.004

14. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., ... Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. Medical Image Analysis, 35, 18–31. https://doi.org/10.1016/j.media.2016.05.004

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 48, pp. 1050–1059).

16. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In Advances in Neural Information Processing Systems 30 (pp. 5574–5584).

17. Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Transactions on Medical Imaging, 35(5), 1299–1312. https://doi.org/10.1109/TMI.2016.2535302

18. Hesamian, M. H., Jia, W., He, X., & Kennedy, P. J. (2019). Deep learning techniques for medical image segmentation: Achievements and challenges. Journal of Digital Imaging, 32(4), 582–596. https://doi.org/10.1007/s10278-019-00227-x