Self-Supervised Representation Learning for Bone Lesion Localization in Three-Dimensional Musculoskeletal Imaging

Marco L. Brock
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
brock1982@missouri.edu

Abstract

Bone lesion localization in three-dimensional musculoskeletal imaging remains difficult because focal marrow replacement, lytic and sclerotic change, fractures, and degenerative alterations can share intensity patterns while differing in clinical meaning. Supervised segmentation and detection models require dense voxel annotations, yet labels for small, multifocal, or low-contrast lesions are scarce and costly. This manuscript develops a methodological framework for self-supervised representation learning that exploits unlabeled computed tomography and magnetic resonance volumes before lesion-specific supervision is introduced. The proposed design combines anatomy-aware contrastive learning with masked volumetric reconstruction, then attaches a voxel-level localization head and a lesion-level aggregation rule. The representation is trained to preserve fine bone texture, spatial context, and cross-scan invariance while limiting shortcuts caused by scanner intensity, patient positioning, and normal anatomical variation. Localization is formulated as a heatmap prediction problem with a binary voxel objective, a noise-robust lesion-level score, and an entropy-based uncertainty map that identifies ambiguous boundary regions. The framework does not assume that self-supervision alone supplies diagnostic semantics; instead, it defines how unlabeled volumes constrain the encoder and how sparse labels refine the decision boundary. Analytical discussion focuses on failure modes, including metal artifact, partial volume effects, marrow heterogeneity, false positives at degenerative margins, and domain shift across acquisition protocols. The contribution is a technically specified and falsifiable training and evaluation logic that connects representation learning to clinically meaningful localization endpoints, including lesion-level sensitivity, false positives per volume, spatial overlap, and calibration, without claiming unsupported empirical performance.

Keywords

self-supervised learning, bone lesion localization, musculoskeletal imaging, three-dimensional representation learning, masked volume modeling, medical image analysis

1. Introduction

Deep learning has become a central methodology for medical image analysis because convolutional and attention-based networks can learn hierarchical features directly from volumetric data [1]. In musculoskeletal imaging, the clinical task is not merely to segment bone or marrow, but to identify focal abnormalities and report their location with sufficient precision to support diagnosis, staging, biopsy planning, and treatment response assessment. Segmentation architectures such as U-Net [2] and self-configuring pipelines such as nnU-Net [3] provide strong building blocks for anatomical delineation, yet they depend on

representative labeled data. Bone lesions are often small, multifocal, and embedded in heterogeneous marrow, and their boundaries may be ambiguous even to expert readers. Dense voxel labels are therefore expensive, inconsistent across observers, and frequently unavailable at the scale required for robust supervised learning.

Self-supervised learning offers a way to use unlabeled volumes as a training signal before lesion annotations are introduced. Contrastive approaches such as SimCLR [4] and momentum contrast [5] learn invariant representations by pulling augmented views together and pushing unrelated examples apart. Bootstrapped latent prediction [6] and self-distillation with vision transformers [7] reduce dependence on explicit negative pairs, while masked image modeling [8] learns by reconstructing hidden content from visible context. These methods are attractive for three-dimensional musculoskeletal data because unlabeled computed tomography and magnetic resonance volumes are abundant relative to lesion-level labels. The central problem is not whether self-supervision can learn generic texture, but whether it can learn representations that preserve the anatomical and pathological distinctions required for localization. A representation that clusters scans by scanner model or body habitus may perform well on common pretext tasks while failing to separate a lytic metastasis from a benign marrow island.

This manuscript addresses that gap by formulating a self-supervised representation learning framework specifically for bone lesion localization in three-dimensional musculoskeletal imaging. The central contribution is a coupled design in which anatomy-aware contrastive learning and masked volumetric reconstruction shape an encoder, while a localization head and lesion-level aggregation rule convert voxel representations into clinically interpretable candidate regions. The framework is developed as a methodological and analytical study rather than an empirical benchmark. Its purpose is to specify the modeling assumptions, objective functions, inference logic, uncertainty treatment, and evaluation endpoints that would make such a system technically coherent and falsifiable. Particular attention is given to the mechanisms by which self-supervision can help or harm lesion localization: preservation of high-frequency bone texture, separation of true lesions from degenerative mimics, robustness to anisotropic voxels, and sensitivity to acquisition-dependent intensity shifts. The resulting argument is that self-supervised representations are most useful for bone lesion localization when they are constrained by anatomical context and when their evaluation is tied to lesion-level detection rather than reconstruction error alone.

2. Problem Formulation and Imaging Constraints

A three-dimensional musculoskeletal examination can be represented as a volume V with spatial dimensions D, H, and W. The localization task is to produce a map S whose value at each voxel reflects the probability that the voxel belongs to a bone lesion or to a lesion-bearing region. In practice, a lesion may be a lytic focus in computed tomography, a marrow-replacing region in magnetic resonance imaging, or a mixed sclerotic and lytic abnormality. The target is therefore not a single organ boundary but a set of spatially distributed, variable-sized, and sometimes low-contrast abnormalities. The mapping can be written as

$$S = f\theta(V), V \in R\hat{}(D \times H \times W)$$

where fθ denotes the full localization network parameterized by θ, and S has the same spatial extent as V. The range of S is the unit interval, with values near one indicating lesion presence. This formulation is intentionally general enough to cover voxel-wise detection, heatmap regression, and candidate-region scoring.

The imaging constraints are decisive. Bone lesions in computed tomography are often defined by cortical destruction or medullary density change, but the same intensity range can be produced by osteophytes, fracture callus, islands of dense bone, or partial volume averaging at vertebral endplates. In magnetic resonance imaging, marrow replacement may appear as altered signal on T1-weighted, fluid-sensitive, or contrast-enhanced sequences, yet red marrow, edema, and treatment-related change can mimic disease. The representation must therefore encode both local texture and broader anatomical context. A patch containing only a lesion margin may be indistinguishable from normal trabecular bone without knowledge of its position relative to the vertebral body, pedicle, endplate, or adjacent disc. This motivates self-supervised objectives that preserve spatial relationships rather than treating each patch as an isolated image.

Three-dimensional self-supervised methods for medical imaging have already shown that volumetric pretext tasks can learn transferable features from scarce labeled data [9]. Transformations such as cube rearrangement and rotation exploit the spatial regularity of anatomy and can be implemented without manual annotation [10]. However, generic three-dimensional pretext tasks may overemphasize global shape and ignore the high-frequency texture that distinguishes a small lytic lesion from normal marrow heterogeneity. A second constraint is anisotropy. Clinical volumes often have different in-plane and through-plane resolution, and aggressive resampling can blur small lesions or introduce interpolation artifacts. The encoder should therefore be trained with augmentation policies that preserve lesion contrast while exposing the network to realistic slice thickness, noise, and reconstruction kernels.

Anatomical segmentation can provide a useful intermediate coordinate system. Transformer-based three-dimensional segmentation models such as UNETR [11] demonstrate that long-range attention can capture organ and tissue context, and robust anatomical segmentation tools such as TotalSegmentator [12] can provide body-region and bone labels for computed tomography. These labels are not lesion labels, but they can define anatomical zones within which lesion candidates are evaluated. Automated detection and localization of vertebral fractures further shows that clinically meaningful bone abnormalities can be approached through structured spatial reasoning rather than whole-volume classification alone [13]. The present framework uses such anatomical context as a prior, not as a substitute for lesion supervision. The localization head must still learn lesion-specific evidence from the limited labeled cases.

3. Self-Supervised Representation Learning Framework

3.1 Volumetric encoding and anatomical context

The encoder Eθ maps an input patch p_i extracted from a volume to a dense feature representation. A projection head gφ maps the pooled representation to a lower-dimensional embedding z_i used for contrastive learning. This relationship is written as

$$z_i = g\varphi\big(E\theta(p_i)\big)$$

where p_i is the i-th patch, Eθ is the shared three-dimensional encoder, φ parameterizes the projection head, and z_i is the resulting embedding. The patch may span a vertebral body, a pelvic region, or a long-bone segment depending on the anatomical zone. The encoder should use anisotropic convolutions or resolution-aware attention so that through-plane blur is not mistaken for lesion absence. The projection head is used only during self-supervised

pretraining; the localization head is attached to the encoder features after pretraining and refined with lesion annotations.

Anatomical context is introduced through patch sampling and positional encoding. Positive pairs are not formed by arbitrary crops from the same volume. Instead, patches are paired when they share an anatomical zone, a similar location relative to a segmented bone structure, or a plausible transformation of the same underlying tissue. This reduces false negatives caused by treating two patches from different patients but the same vertebral level as unrelated. The design also reduces the risk that the encoder solves the pretext task by memorizing patient-specific intensity profiles. Positional encodings are normalized within the anatomical zone so that the model can distinguish a lesion near an endplate from one in the pedicle without learning absolute scanner coordinates.

3.2 Contrastive and masked reconstruction objectives

The contrastive objective encourages agreement between different views of the same anatomical content while separating unrelated content. For an anchor embedding z_i, a positive embedding z_i^+, and N embeddings in the comparison set, the loss is

$$L_con = -\log\left(\exp\frac{\left(sim\frac{(z_i, z_i^+)}{\tau}\right)}{\Sigma_{j=1}^{N}}\exp\left(sim\frac{(z_i, z_j)}{\tau}\right)\right)$$

where sim denotes cosine similarity, τ is a positive temperature parameter, and the sum runs over positive and negative embeddings in the comparison set. The positive view is generated by intensity windowing, mild elastic deformation, noise injection, and small affine changes that preserve lesion appearance. Strong geometric transformations are avoided because a rotated or heavily warped lesion may no longer occupy the same anatomical relation to the cortex or endplate. The comparison set should include hard negatives from the same anatomical zone, because negatives drawn only from distant body regions allow the encoder to separate coarse anatomy without learning subtle lesion texture.

Masked volumetric reconstruction complements contrastive learning by forcing the encoder to infer missing content from surrounding bone and marrow. Let M be a binary mask indicating voxels that are hidden, and let rψ be a decoder. The reconstruction loss is

$$L_rec = \Sigma_\{m \in M\} ||r\psi(E\theta(V \odot (1 - M)))_m - V_m||^2$$

where ⊙ denotes element-wise multiplication, ψ parameterizes the decoder, and the sum is over masked voxels. A high masking ratio encourages the encoder to use global context, while a moderate ratio preserves enough local texture for fine lesion boundaries. The reconstruction target should be computed in a normalized bone window or a modality-specific intensity space so that the loss is not dominated by fat, air, or metal artifacts. This objective resembles contrastive pretraining in other medical domains in that it learns transferable features from unlabeled images [15], but the sampling and masking policy here is tied to musculoskeletal anatomy rather than generic image patches.

The two objectives serve different roles. Contrastive learning shapes a discriminative embedding space and can improve robustness to acquisition variation. Masked reconstruction preserves local structure and forces the encoder to represent fine-grained texture. Using both is appropriate because bone lesion localization requires both global anatomical reasoning and

local boundary evidence. The risk is that reconstruction may devote capacity to normal anatomy, which is abundant, while rare lesions contribute little to the loss. This risk is mitigated by anatomy-aware masking that targets bone and marrow regions and by contrastive sampling that includes subtle negative examples such as degenerative margins and treated lesions.

3.3 Localization head and weakly supervised refinement

After self-supervised pretraining, the encoder is connected to a localization head hv that predicts a voxel-level probability map. The head can be a lightweight three-dimensional convolutional decoder or a dense prediction transformer that upsamples features to the original resolution. The training objective for labeled voxels is a binary cross-entropy loss over the set Ω of annotated voxels:

$$L_loc = -\Sigma_{v\in\Omega}\left[Y(v)\log S(v) + \left(1 - Y(v)\right)\log\left(1 - S(v)\right)\right]$$

where Y(v) ∈ {0,1} is the lesion label at voxel v and S(v) is the predicted probability. In many datasets, only a subset of slices or lesion centroids is annotated. In that case, Y(v) can be defined on a dilated lesion mask or on a Gaussian-smoothed target around the centroid, with the understanding that centroid supervision evaluates localization rather than exact boundary segmentation. The head should be trained with hard-negative mining around fracture lines, osteophytes, and endplate irregularities, because these structures are common sources of false positives.

A related line of work uses segmentation masks to guide semi-supervised landmark detection in three-dimensional scans, which supports the use of anatomical priors when labels are sparse [14]. In the present framework, segmentation masks are not treated as lesion labels. They define anatomical zones, constrain candidate locations, and provide auxiliary boundaries that help the localization head distinguish lesion tissue from adjacent normal structures. This use of segmentation guidance is consistent with the broader principle that self-supervised and semi-supervised signals should shape the representation without inventing diagnostic semantics that the data do not contain.

3.4 Combined objective and uncertainty

The self-supervised and localization terms are combined during fine-tuning as

$$L = \lambda_1 L_rec + \lambda_2 L_con + \lambda_3 L_loc + \lambda_4 ||\theta||_2^2$$

where λ_1, λ_2, λ_3, and λ_4 are nonnegative weights, and $||\theta||_2^2$ is a weight decay term. The weights express a trade-off rather than a fixed prescription. A large λ_1 keeps the encoder faithful to local texture but may slow adaptation to lesion-specific features. A large λ_2 improves invariance but can suppress subtle intensity differences if augmentations are too aggressive. A large λ_3 prioritizes the limited labels but risks overfitting to a small set of lesion appearances. The combined objective therefore makes the design assumptions explicit and allows sensitivity to be analyzed without claiming a particular setting is optimal.

Uncertainty should be reported alongside the probability map. For a voxel probability S(v), the predictive entropy is

$$U(v) = -S(v)\log S(v) - \left(1 - S(v)\right)\log\left(1 - S(v)\right)$$

where U(v) is maximal when S(v) is near 0.5 and minimal when the prediction is confident. Entropy does not by itself provide calibrated uncertainty, but it is a useful

screening signal for ambiguous boundaries, metal artifact, and partial volume regions. Bayesian approximation methods such as Monte Carlo dropout [16] and approximate Bayesian inference in deep networks [17] can provide a more principled uncertainty estimate if multiple stochastic forward passes are available. In either case, uncertainty should be evaluated as a localization aid, not as a substitute for lesion-level validation.

4. Localization Inference and Evaluation Logic

At inference, the voxel probability map is converted into candidate lesions by thresholding, connected-component analysis, and anatomical zone filtering. For a candidate region R_l, a lesion-level score can be defined by a noisy-OR aggregation:

$$P_l = 1 - \prod_{v \in R_l} \big(1 - S(v)\big)$$

where P_l is the probability that region R_l contains at least one lesion and the product runs over voxels in the region. This rule assumes conditional independence of voxel evidence given the region, an assumption that is clearly approximate because neighboring voxels are correlated. Its advantage is interpretability: a region with several moderately suspicious voxels can receive a high score even if no single voxel is highly confident. Alternative aggregation rules, such as top-k averaging or learned region classifiers, should be compared against this baseline because the choice affects sensitivity to small lesions and false-positive burden.

Evaluation must separate voxel-level accuracy from lesion-level utility. A model can achieve high voxel overlap on large lesions while missing small lesions entirely, and a model can detect every lesion while producing an unacceptable number of false positives. Detection benchmarks in computed tomography have established lesion-level sensitivity and false positives per scan as informative endpoints [18]. For voxel-level outputs, overlap metrics such as the Dice coefficient [19] and their three-dimensional properties [20] provide complementary information about boundary agreement. Localization error should also be reported as the distance between the predicted region centroid and the annotated lesion centroid, because a clinically useful system must identify the correct anatomical location even when the boundary is imperfect. Calibration of P_l and S(v) should be assessed with reliability diagrams or proper scoring rules so that a high score corresponds to a meaningful probability rather than an arbitrary ranking.

The evaluation design should account for lesion size, anatomical site, modality, and acquisition protocol. Small lesions, lesions near metal hardware, and lesions in regions with complex marrow distribution are likely to produce different error profiles. Stratified reporting is therefore more informative than a single aggregate score. If multiple readers annotate the same cases, inter-observer agreement should be reported because the target itself is uncertain at boundaries. Finally, external validation across scanners and institutions is essential for any claim of clinical transportability. The present framework specifies these endpoints without asserting that they have been achieved.

5. Technical Discussion

The proposed framework has several strengths that follow from its structure. First, self-supervised pretraining can use unlabeled volumes to learn bone texture and anatomical context before lesion labels are introduced. This is particularly relevant in musculoskeletal imaging, where unlabeled computed tomography and magnetic resonance studies are common

but lesion-level annotations are sparse. Second, the combination of contrastive learning and masked reconstruction addresses two different failure modes. Contrastive learning improves invariance to acquisition variation, while reconstruction preserves fine-grained structure. Third, the localization head and lesion-level aggregation rule connect representation learning to a clinically meaningful output rather than leaving the model as a generic feature extractor.

The framework also has important limitations. Self-supervision can learn shortcuts that are invisible in reconstruction or contrastive loss. For example, the encoder may cluster volumes by scanner manufacturer, slice thickness, or contrast phase. If those factors correlate with lesion prevalence in the fine-tuning data, the model can appear accurate for the wrong reason. Anatomy-aware sampling and stratified evaluation reduce this risk but do not eliminate it. A second limitation is that bone lesions are rare relative to normal bone voxels. Contrastive learning may treat subtle lesions as negatives because they are statistically uncommon, and reconstruction may reconstruct them poorly because the model has learned normal anatomy. Hard-negative sampling and lesion-focused masking are intended to counter this, but the balance between normal and abnormal content remains a design choice with no universal solution.

A third limitation concerns annotation geometry. Lesion boundaries are often ill-defined, especially for infiltrative marrow disease or treated lesions with sclerosis. A model trained on centroid labels may learn to predict a blob around the lesion rather than the true extent, while a model trained on dense masks may inherit observer-specific boundary conventions. The evaluation should therefore report both centroid-based localization error and boundary overlap, and it should acknowledge that neither is a pure measure of clinical truth. A fourth limitation is domain shift. Intensity distributions, reconstruction kernels, and patient positioning vary across scanners and protocols. Self-supervised representations may be more robust than supervised ones trained on small datasets, but robustness must be demonstrated rather than assumed.

The framework's practical significance lies in its testable structure. Each component corresponds to an observable quantity: contrastive loss, reconstruction error, localization loss, lesion-level score, entropy, overlap, and calibration. This makes the design amenable to ablation and error analysis. For instance, removing anatomy-aware sampling should increase false positives at degenerative margins if the hypothesis is correct. Replacing masked reconstruction with a generic denoising objective should affect small-lesion sensitivity if high-frequency texture is important. Such analyses would be more informative than reporting a single aggregate score, because they would identify which self-supervised mechanism contributes to which aspect of localization.

Clinically, the output should be framed as a localization aid rather than an autonomous diagnosis. A heatmap that highlights a suspicious marrow region can support a radiologist's search pattern, but it cannot establish the cause of the abnormality. The system should therefore report candidate regions with uncertainty and anatomical labels, allowing the reader to distinguish a high-confidence lesion from an ambiguous area near hardware or an endplate. This distinction is especially important in oncology, where bone lesions may be multifocal and where missing a small lesion can alter staging. The technical challenge is to preserve sensitivity without overwhelming the reader with false positives, and that trade-off must be evaluated in the intended clinical workflow.

6. Conclusion

This manuscript has developed a self-supervised representation learning framework for bone lesion localization in three-dimensional musculoskeletal imaging. The framework combines anatomy-aware contrastive learning with masked volumetric reconstruction, attaches a voxel-level localization head, and aggregates voxel evidence into lesion-level scores with an explicit uncertainty map. The central argument is that self-supervision is useful for this task only when it is constrained by musculoskeletal anatomy and evaluated at the level of lesion detection rather than reconstruction alone. The proposed objectives make the assumptions visible: positive pairs reflect anatomical correspondence, augmentations preserve lesion appearance, masking targets bone and marrow, and the localization loss refines the representation with sparse labels. The discussion identifies failure modes that are specific to bone imaging, including degenerative mimics, partial volume effects, metal artifact, marrow heterogeneity, and scanner-dependent domain shift. The framework does not claim empirical superiority over existing systems. It provides a technically coherent and falsifiable basis for developing and assessing self-supervised representations, with evaluation anchored in lesion-level sensitivity, false positives per volume, spatial overlap, localization error, and calibration.

References

1. 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

2. 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

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. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning (PMLR 119, pp. 1597–1607).

5. He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9729–9738). IEEE. https://doi.org/10.1109/CVPR42600.2020.00975

6. Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z. D., Gheshlaghi Azar, M., Piot, B., Kavukcuoglu, K., Munos, R., & Valko, M. (2020). Bootstrap your own latent: A new approach to self-supervised learning. In Advances in Neural Information Processing Systems (Vol. 33, pp. 21271–21284).

7. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9650–9660). IEEE. https://doi.org/10.1109/ICCV48922.2021.00951

8. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16000–16009). IEEE. https://doi.org/10.1109/CVPR52688.2022.01553

9. Taleb, A., Loetzsch, W., Danz, N., Severin, J., Gaertner, T., Bergner, B., & Lippert, C. (2020). 3D self-supervised methods for medical imaging. In Advances in Neural Information Processing Systems (Vol. 33, pp. 18158–18172).

10. Zhuang, X., Li, Y., Hu, Y., Ma, K., Yang, Y., & Zheng, Y. (2019). Self-supervised feature learning for 3D medical images by playing a Rubik's cube. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2019 (pp. 420–428). Springer.

11. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., & Xu, D. (2022). UNETR: Transformers for 3D medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 574–584). IEEE.

12. Wasserthal, J., Breit, H.-C., Meyer, M. T., Prado, M., & Sauter, D. (2023). TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. Radiology: Artificial Intelligence, 5(5), e230024.

13. Burns, J. E., Yao, J., Wiese, T. S., Muñoz, H. E., & Summers, R. M. (2017). Automated detection, localization, and classification of traumatic vertebral body fractures in the thoracolumbar spine. Journal of Neurosurgery: Spine, 27(6), 675–684.

14. 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

15. Ciga, O., Xu, T., & Martel, A. L. (2022). Self supervised contrastive learning for digital histopathology. Machine Learning with Applications, 7, 100198.

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 (Vol. 30, pp. 5574–5584).

17. 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).

18. Setio, A. A. A., Traverso, A., de Bel, T., Berens, M. S., van den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M. E., Geurts, B., Gugten, R. van, Ho, H., Jansen, B., Kasteleijn, A., Korez, R., Lassen, B., López, M., Lozano, R., Mi, H., ... Jacobs, C. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Medical Image Analysis, 42, 1–13.

19. Dice, L. R. (1945). Measures of the amount of ecologic association between species. Ecology, 26(3), 297–302.

20. Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Medical Imaging, 15, 29.