IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 May 2026, accepted 24 May 2026, date of publication 1 June 2026, date of current version 8 June 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3698609

RESEARCH ARTICLE

Anomaly and Spatially Aware Pretraining for Brain Hemorrhage Segmentation in Limited Annotated Data Setting

MARYNA KVASNYTSIA[1,2], ABEL DÍAZ BERENGUER[1,2], HICHEM SAHLI[1,2], AND JEF VANDEMEULEBROUCKE[1,2,3]

[1]Department of Electronics and Informatics, Vrije Universiteit Brussel, 1050 Brussels, Belgium
[2]IMEC, 3001 Leuven, Belgium
[3]Department of Radiology, UZ Brussel, Vrije Universiteit Brussel, 1090 Brussels, Belgium

Corresponding author: Maryna Kvasnytsia (maryna.kvasnytsia@vub.be)

This work was supported by the Interuniversity Microelectronics Centre (imec) Interdisciplinary Cooperative Research (ICON) project "Label-efficient medical image analysis" (LEMMA), funded by the Flemish Agency for Innovation and Entrepreneurship (VLAIO), under Grant HBC.2024.1016.

ABSTRACT Lesion segmentation in medical imaging is challenging due to limited voxel-level annotations and high variability in pathological appearance. Although self-supervised learning (SSL) enables the use of unlabeled data, most contrastive approaches do not explicitly incorporate pathology-related information, and reconstruction-based methods often remain more effective for downstream segmentation tasks. We propose an anomaly-modulated spatially-aware contrastive learning framework (AMSA) that integrates anatomical localization with anomaly-derived signals to improve representation learning in low-label regimes. First, images are affinely registered and decomposed into spatially aligned patches. Next, approximate anomaly maps are obtained via an image-to-image translation model and summarized into continuous anomaly scores. Spatial distance between patches is used to weigh interactions in the contrastive objective. Additionally, the contrastive objective is modulated with an estimated anomaly burden, encouraging representations to remain anatomically consistent while incorporating pathology-related information into the representation space. The pretrained encoder is subsequently fine-tuned for segmentation. Experiments on brain CT hemorrhage segmentation show consistent improvements over fully supervised training and spatially aware contrastive pretraining, while achieving performance comparable to reconstruction-based pretraining methods.

INDEX TERMS Anomaly modulation, brain hemorrhage, lesion segmentation, spatial prior, self-supervised learning, contrastive learning.

I. INTRODUCTION

Detection and segmentation of intracranial hemorrhage (ICH) in computed tomography (CT) are clinically critical, as hemorrhage volume and spatial distribution directly inform prognosis, treatment planning, and monitoring of disease progression [1], [2], [3], [4], [5]. However, automated ICH segmentation remains challenging due to substantial variability in lesion size, shape, and location. Hemorrhages may appear as small focal bleeds or diffuse patterns, and often occupy only a small fraction of the image volume. In addition, early or mixed-density hemorrhages can exhibit subtle intensity differences relative to the surrounding tissue, making them difficult to distinguish from normal anatomical variation. These characteristics result in a highly imbalanced and heterogeneous learning problem, in which small pathological regions must be detected against the dominant healthy anatomy.

Deep learning methods for medical image analysis, including segmentation, detection, and classification, have achieved remarkable success. However, their performance typically relies on large amounts of high-quality annotated data, the acquisition of which often requires expert knowledge. In practice, voxel-level annotations are expensive and

The associate editor coordinating the review of this manuscript and approving it for publication was Majdi Mansouri.

time-consuming to obtain, particularly for small and ambiguous lesions such as microbleeds. Therefore, self-supervised learning (SSL) has emerged as a promising paradigm for learning transferable image representations from large collections of unlabeled medical data. Despite recent progress, designing effective SSL methods for medical imaging faces significant challenges [6], [7]. In contrast to natural images, medical images exhibit subtle intensity variations, strong anatomical structure, and task-specific semantic cues. These challenges are particularly evident in lesion segmentation tasks, where target structures are often small, heterogeneous, and sparsely distributed. Generic SSL objectives that encourage global invariance or image-level discrimination may therefore fail to capture features that are relevant for downstream lesion segmentation, instead focusing predominantly on normal anatomical variability, as illustrated in Figure 1.

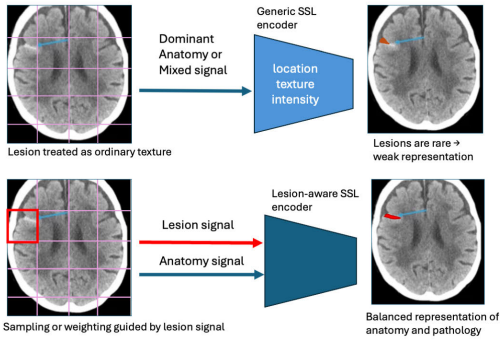


FIGURE 1. Conceptual illustration of generic versus lesion-aware self-supervised pretraining. Generic SSL methods learn representations primarily driven by anatomical structure because pathological regions are sparse and treated as ordinary image content. As a result, lesion-specific features may be weakly represented. In contrast, lesion-aware SSL incorporates signals related to anomaly or lesion presence during pretraining, enabling the model to learn representations sensitive to both anatomical and pathological variation, which may improve performance on downstream lesion segmentation tasks.

Considering the tasks of ICH detection and segmentation, we suggest that from a representation-learning perspective, effective pretraining should simultaneously preserve anatomical coherence and enhance sensitivity to subtle pathological deviations. In many cases, pathology present itself as localized lesions and could be seen as anomalous regions within a repetitive anatomical context. Anomaly detection approaches could therefore provide an approximate localization of pathological tissue. Therefore, we hypothesize that guiding self-supervised pretraining with anomaly-derived signals while constraining interactions to anatomically corresponding regions can encourage representations towards pathology-relevant features without sacrificing spatial consistency. This hypothesis motivated the proposed anomaly-modulated spatially aware contrastive learning framework.

The contributions of this study are as follows. We propose a contrastive learning framework that leverages noisy anomaly maps obtained from weakly-supervised image-to-image translation models to softly emphasize lesion-related regions during pretraining. Rather than treating all image regions equally, our approach encourages representation learning towards anatomically corresponding healthy regions while modulating similarity according to the estimated anomaly burden. We show that anomaly-aware modulation provides complementary benefits to spatial contrastive pretraining for downstream lesion detection and segmentation in low-label regimes. Current study provides a conceptual perspective on how pathology-aware signals can be integrated into spatially structured self-supervised learning.

II. RELATED WORK AND CONTRIBUTIONS

Early SSL approaches in medical imaging primarily relied on generic pretext tasks, such as image restoration, context prediction, or transformation recovery [8], [9], [10]. Model Genesis [11] is a representative example that combines multiple restoration-based objectives to encourage the learning of anatomical structures. While effective, such methods treat normal and abnormal tissues uniformly during pretraining and do not explicitly account for the presence or spatial distribution of pathology.

More recent SSL methods have incorporated spatial or anatomical priors into contrastive learning objectives. Chaitanya et al. [12] proposed local contrastive learning by contrasting features extracted from the corresponding spatial locations across augmented views, thereby enforcing spatial consistency.Voxel or region-level contrastive objectives have been introduced to preserve the anatomical structure in volumetric data [13]. Although these approaches improve the anatomical coherence and spatial alignment of learned representations, they remain agnostic to lesion presence and do not exploit abnormality cues during representation learning.

Several studies have attempted to introduce lesion or abnormality information into the representation learning. Most such approaches rely on explicit lesion annotations, region-of-interest masks, or coarse abnormality labels to guide pretraining, effectively introducing strong-to-moderate supervision [14], [15]. For instance, lesion-aware contrastive learning has been explored in both histopathology and radiology by constraining feature similarity within annotated lesion regions [16], [17]. Although these methods can improve downstream performance, their applicability is limited by the availability, quality, and consistency of lesion annotations, making them unsuitable for fully unlabeled pretraining scenarios.

An alternative and complementary line of research focuses on anomaly detection, where SSL is used to model the normal anatomy and detect deviations from it [18], [19]. Recent methods incorporate representation learning more directly: anomaly signals can be used to guide feature learning, improve latent separability, or integrate semantic priors. For instance, [20] integrated self-supervised semantic

priors into generative modeling, whereas [21] directly shaped latent representations using anomaly decision boundaries. Other approaches exploit discrepancy learning between normal and mixed data [22], anomaly-inspired prototype distances [23], or anatomical priors via prompt learning [24]. However, these methods typically use anomaly signals for detection or scoring, rather than embedding them directly into the representation learning objectives. To the best of our knowledge, previous studies have not leveraged anomaly maps as a continuous modulation signal within a contrastive loss formulation.

III. ANOMALY-MODULATED SPATIALLY-AWARE PRETRAINING

We propose an anomaly-modulated spatially-aware contrastive pretraining framework for lesion segmentation. A schematic representation of the proposed framework is shown in Figure 2.

The images were first affinely registered to a common template and partitioned into spatially aligned patches using a fixed anatomical grid. In parallel, anomaly maps are generated offline using an image-to-image translation model trained with image-level supervision and summarized into a continuous anomaly burden signal [25]. Anomaly burden is estimated as the fraction of voxels with intermediate anomaly responses, providing a stable measure of anomaly extent while reducing the influence of noise and outliers. During self-supervised pretraining, an encoder is optimized using the proposed anomaly-modulated spatial contrastive loss (AMSA).

The proposed loss function is designed to induce a representation space structured primarily by anatomical locality, while incorporating pathology as a secondary, softly modulating factor. Specifically, patches originating from the same anatomical location are encouraged to form coherent clusters, whereas similarity between patches from different locations decreases as a function of spatial distance. This results in a representation space that is smoothly organized according to anatomy rather than being discretely partitioned.

Within each anatomical region, anomaly modulation introduces a secondary effect by adjusting the similarity between patches based on the estimated pathology burden. Specifically, anomaly-derived scores are used to modulate pairwise interaction weights within the spatially weighted contrastive objective. Rather than enforcing explicit separation between healthy and pathological samples, this mechanism encourages the soft stratification of representations, reflecting varying degrees of abnormality. Given the heterogeneous nature of lesions, including variability in size, shape, and appearance, strongly separated subclusters are not expected. Instead, the model leverages anatomical structure as the primary organizing principle, while anomaly information acts as a complementary signal that refines representation without disrupting spatial coherence. The pretrained encoder is subsequently fine-tuned with a segmentation decoder using limited labeled data.

In the following sections, we describe each component of the pipeline in detail.

A. SPATIAL ALIGNMENT AND PATCH EXTRACTION

To ensure spatial coherence across the subjects, all volumetric images are first aligned to a common anatomical reference. Specifically, each input volume is affinely registered to a fixed template using a global affine transformation. This preprocessing step standardized orientation, scale, and global anatomy, enabling consistent correspondence between spatial locations across subjects.

After registration, a fixed grid of spatial anchor points was defined in the template space. At each anchor location, 3D patches of fixed size were extracted from all subjects. Therefore, each patch is associated with a discrete anatomical location label ℓ and a corresponding spatial coordinate $\mathbf{c}_\ell$ in the template space. This design ensures that patches with the same location label originate from anatomically equivalent regions across the dataset, which is a prerequisite for spatially grounded contrastive learning.

B. CONTRASTIVE ENCODER ARCHITECTURE CHOICE

We adopted a deliberately simple encoder architecture to focus on the contribution of the proposed anomaly-modulated contrastive loss. The encoder follows a lightweight 3D U-Net–style design and consists exclusively of a downsampling (encoder) path.

Formally, given an input patch $x \in \mathbb{R}^{H\times W\times L}$, the encoder $f(\cdot)$ is composed of a sequence of four convolutional U-Net blocks where each block comprises 3D convolutions, normalization layers, and nonlinear activations, with spatial downsampling implemented through strided convolutions. The final feature map is aggregated using global average pooling, yielding a fixed-dimensional embedding:

$$\mathbf{z} = \mathrm{GAP}\big(f(x)\big) \in \mathbb{R}^D, \tag{1}$$

which is subsequently L2-normalised before being passed to the contrastive objective.

The encoder architecture is intentionally kept shallow and parameter-efficient to minimize architectural complexity, allowing the effect of the proposed anomaly-modulated contrastive loss to be isolated and analyzed without confounding factors. Second, it demonstrates that meaningful spatially and pathology-aware representations can be learned even with a simple backbone. The proposed loss function is independent of the encoder architecture. In practice, representation quality can be further improved by replacing the encoder with deeper or more expressive backbones, such as deeper U-Net variants, residual networks, or transformer-based encoders. In this study, however, we employ a lightweight U-Net-style encoder as a proof-of-concept to focus on the core contribution: anomaly-modulated spatial contrastive learning.

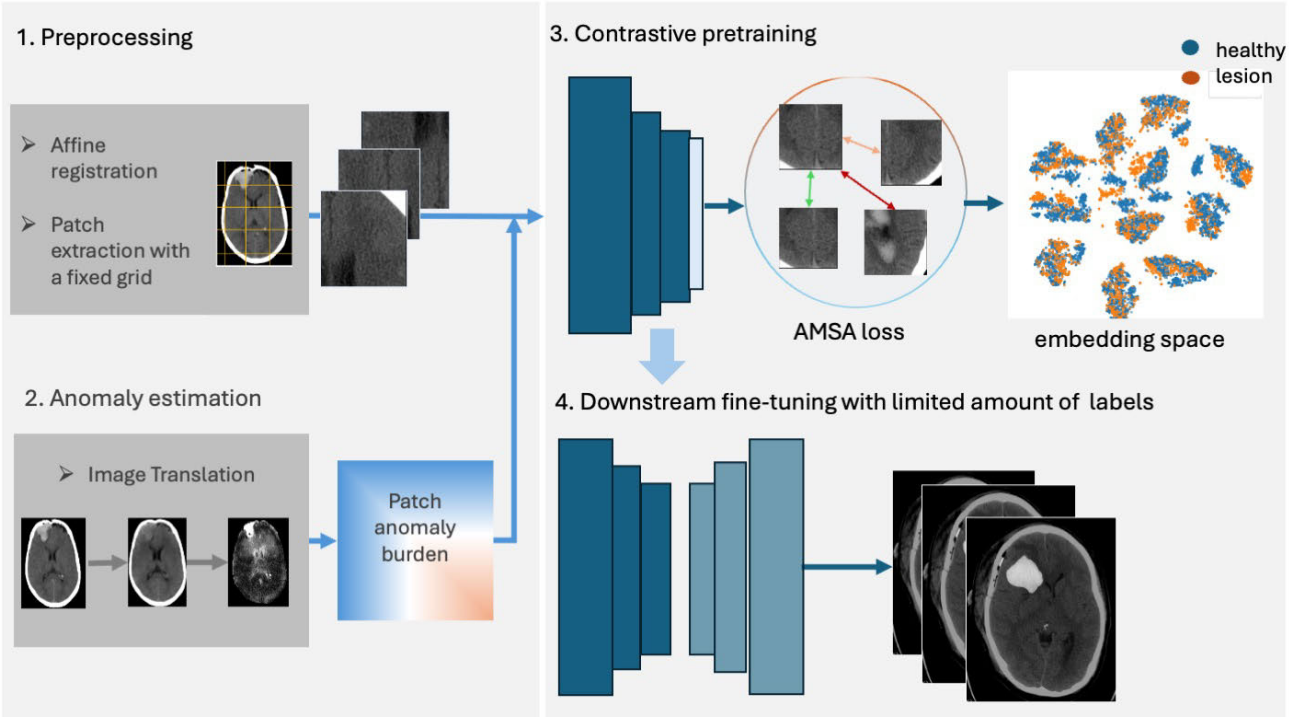


FIGURE 2. Overview of the proposed framework. All images are first affinely registered to a common template and decomposed into spatially aligned patches using a fixed anatomical grid. In parallel, anomaly maps are obtained offline using an image-to-image translation model trained with image-level supervision, and summarized as a continuous anomaly burden signal. During contrastive pretraining, a shared encoder is optimized using the proposed anomaly-modulated spatial contrastive loss (AMSA), which enforces locality-aware positive pairs while modulating interactions based on anomaly differences. The pretrained encoder is subsequently fine-tuned with a segmentation decoder using limited labeled data.

C. ANOMALY MAP GENERATION VIA UNPAIRED IMAGE TRANSLATION

To obtain anomaly maps without relying on lesion annotations, an unsupervised image-to-image translation approach was employed. Specifically, we train a Contrastive Unpaired Image Translation (CUT [25]) model to translate images from the pathological domain to a healthy-appearing domain.

The CUT framework learns mapping $G : \mathcal{X}_{\text{path}} \rightarrow \mathcal{X}_{\text{healthy}}$ using unpaired data, enforcing content preservation through patch-level contrastive learning while adapting the image appearance to the target domain. Consequently, the generator suppresses the pathological patterns that are inconsistent with healthy anatomy, while largely preserving normal tissue structure.

Given a pathological input image x, anomaly map a is computed as the voxel-wise absolute difference between the input and its translated counterpart:

$$a = |x - G(x)|. \tag{2}$$

This difference image highlights regions that are altered by the translation process and therefore serves as a proxy for abnormal or pathological content.

The resulting anomaly maps are inherently noisy and do not correspond to ground-truth lesion segmentations. These anomaly maps are not suitable as supervision signals. Instead, they provide a weak estimate of the relative anomaly burden, which is subsequently aggregated into a scalar anomaly score and used to modulate contrastive loss. This design allows the proposed method to incorporate pathology-related information while remaining fully self-supervised with respect to the lesion annotations.

D. AMSA LOSS FORMULATION

We propose a spatially and lesion-aware contrastive objective based on a weighted NT-Xent formulation [26]. Positive pairs were defined as healthy patches originating from the same anatomical location, whereas all other pairs acted as negatives. Pairwise contributions are modulated by an exponential spatial decay and a lesion contrast term derived from the anomaly load estimates. This design enforces anatomical consistency while sensitizing representations to pathological severity without collapsing lesion samples into a single cluster.

Instead of relying on ground-truth lesion annotations, we derive a proxy lesion load from anomaly maps generated by an independent unsupervised model. These anomaly maps are inherently noisy and may contain artifacts unrelated to pathology. To ensure robustness, we aggregated anomaly responses using a robust operator and normalized them across the batch. The resulting scalar reflects the relative anomaly burden rather than true lesion volume. This proxy is used to modulate the contrastive loss, down-weighting pairs with a

disparate anomaly burden while preserving the anatomical structure. As a result, the proposed loss function remains applicable in settings where lesion labels are unavailable or noisy.

1) ANOMALY-MODULATED SPATIAL CONTRASTIVE LOSS

Let a mini-batch consist of B image patches $\{x_i\}_{i=1}^{B}$. Each patch is associated with: (i) a discrete anatomical location label ℓ_i, (ii) a spatial coordinate $\mathbf{c}_{\ell_i} \in \mathbb{R}^d$ (i.e., the center of grid cell ℓ_i), and (iii) an anomaly map a_i obtained from an independent image translation model.

An encoder $f(\cdot)$ maps each patch to a feature vector, that is L2-normalized:

$$\mathbf{z}_i = \frac{f(x_i)}{\|f(x_i)\|} \in \mathbb{R}^D. \tag{3}$$

From each anomaly map a_i, we compute a scalar proxy for lesion burden

$$\tilde{r}_i = \mathcal{R}(a_i), \tag{4}$$

where $\mathcal{R}(\cdot)$ denotes a robust aggregation operator that estimates the spatial extent of anomalous signal. Specifically, $\mathcal{R}$ measures the fraction of voxels whose anomaly responses fall within a predefined mid-range interval $[\tau_{\text{low}} = 40, \tau_{\text{high}} = 90]$, thereby suppressing the background noise and extreme outliers commonly present in anomaly maps. This yields a stable, bounded estimate of the anomaly extent rather than peak intensity. The resulting values were normalized across the batch using a bounded min–max transformation:

$$r_i = \frac{\tilde{r}_i - \min_j \tilde{r}_j}{\max_j \tilde{r}_j - \min_j \tilde{r}_j + \varepsilon}, r_i \in [0, 1]. \tag{5}$$

This normalization produces scale-invariant modulation signals, while preserving the relative differences in anomaly burden within each batch. The additive term $\varepsilon > 0$ prevents numerical instability in cases with limited anomaly variation.

To identify low-anomaly patches, we define a data-adaptive threshold

$$\theta = \max\left(Q_q(\{r_i\}), \theta_{\min}\right), \tag{6}$$

where Q_q denotes the empirical q-quantile (e.g., $q = 0.1$). A patch was considered *healthy* if $r_i \le \theta$.

This adaptive threshold does not define the ground-truth normal or healthy tissue, but instead provides a relative criterion for selecting patches with a low estimated anomaly burden within each batch.

A pair (i, j) with $i \neq j$ is considered positive if both patches originate from the same anatomical location and exhibit a low anomaly burden:

$$\mathbb{1}_{ij}^{+} = \begin{cases} 1, & \text{if } \ell_i = \ell_j \text{ and } r_i \le \theta \text{ and } r_j \le \theta, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

All remaining pairs implicitly contribute as negatives.

We used cosine similarity with temperature scaling, as follows:

$$s_{ij} = \frac{\mathbf{z}_i^{\top} \mathbf{z}_j}{\tau}, \tag{8}$$

where $\tau > 0$ is a temperature parameter. For numerical stability, the similarities were shifted per anchor:

$$\tilde{s}_{ij} = s_{ij} - \max_{k \neq i} s_{ik}. \tag{9}$$

Spatial locality is enforced using exponential decay with respect to the inter-location distance:

$$w_{ij}^{\text{spatial}} = \exp\left(-\frac{\|\mathbf{c}_{\ell_i} - \mathbf{c}_{\ell_j}\|}{\sigma}\right), \tag{10}$$

where σ controls the spatial interaction scale.

Differences in anomaly burden modulate pairwise interactions via

$$w_{ij}^{\text{anom}} = \exp\left(-\beta |r_i - r_j|\right), \tag{11}$$

where $\beta \ge 0$ controls the strength of the anomaly modulation.

Anomaly modulation influences the contrastive objective through the normalization term, driving representation learning toward anomaly-consistent neighborhoods without altering the definition of positive pairs. We apply anomaly weighting only in the normalization term to modulate negative competition while preserving the positive pair alignment.

The final pairwise interaction weight is given by

$$w_{ij} = w_{ij}^{\text{spatial}} \cdot w_{ij}^{\text{anom}}. \tag{12}$$

For each anchor i, we define a weighted partition function

$$Z_i = \sum_{j \neq i} w_{ij} \exp(\tilde{s}_{ij}). \tag{13}$$

The proposed anomaly-modulated spatial contrastive loss is defined as

$$\mathcal{L} = -\frac{\sum_{i=1}^{B} \sum_{j=1}^{B} \mathbb{1}_{ij}^{+} \left(\tilde{s}_{ij} - \log Z_i\right)}{\sum_{i,j} \mathbb{1}_{ij}^{+} + \varepsilon}. \tag{14}$$

2) CHOICE OF CONTRASTIVE OBJECTIVE

We adopted an InfoNCE-style contrastive objective rather than a triplet loss formulation. Triplet losses rely on explicit anchor-positive-negative tuples and require the hard assignment of samples into positive and negative sets. Such formulations are ill-suited to our setting, where pairwise relationships are inherently graded owning to the spatial proximity and continuously valued anomaly estimates derived from noisy
anomaly maps.

The InfoNCE objective naturally incorporates multiple negatives per anchor and allows for continuous pairwise weighting. This property is essential for our anomaly-modulated framework, in which both the spatial locality and anomaly burden are encoded as soft weights rather than binary constraints. Moreover, InfoNCE avoids the need for hard negative mining strategies, which are known to be sensitive to noise and unstable in practice. InfoNCE-style objectives have demonstrated superior stability and scalability in self-supervised learning, particularly in medical imaging applications [27].

IV. LOSS DESIGN PRINCIPLE

A. DESIGN PRINCIPLES FOR SPATIAL AND ANOMALY MODULATION

The proposed contrastive objective is governed by two coupled hyperparameters: the spatial decay parameter σ and anomaly modulation strength β. Their effects are distinct but interdependent. The spatial parameter σ determines where contrastive forces act, whereas β determines how strongly pathology modulates similarity.

1) DESIGN PRINCIPLE 1 (SPATIAL LOCALITY)

Let $\mathbf{c}_\ell \in \mathbb{R}^d$ denote the spatial coordinate associated with anatomical location ℓ, and let

$$\Delta = \text{median}_{\ell \sim \ell'} \|\mathbf{c}_\ell - \mathbf{c}_{\ell'}\| \tag{15}$$

denote the characteristic inter-location spacing of the patch grid. The spatial decay parameter should be proportional to the intrinsic grid spacing.

$$\sigma^\star \approx k\,\Delta, \qquad k \in [0.7,\ 1.2]. \tag{16}$$

When $\sigma \ll \Delta$, contrastive interactions are overly restricted to identical locations, limiting information sharing across adjacent anatomies. When $\sigma \gg \Delta$, spatial weighting becomes nearly uniform, effectively removing anatomical locality and reducing the objective to a global contrastive loss. Empirically, matching σ to the inter-location spacing yields an optimal performance and stability.

2) DESIGN PRINCIPLE 2 (CONDITIONAL ANOMALY MODULATION)

The anomaly modulation parameter β controls the sensitivity of the contrastive objective to differences in anomaly severity. Increasing β improves pathology sensitivity only when the spatial interactions are local. For $\sigma \gg \sigma^\star$, increasing β may introduce competing objectives between global anomaly separation and anatomical consistency, leading to degraded performance. Because the anomaly-derived lesion burden r_i is estimated from noisy anomaly maps, increasing β amplifies both the informative pathology signals and estimation noise. In practice, moderate values of β provide the best trade-off between the lesion sensitivity and training stability.

V. METHODS

A. BASELINES AND COMPARISON PROTOCOLS

To assess the effectiveness of the proposed anomaly-modulated contrastive pretraining, we compare our approach with both fully supervised and self-supervised learning baselines under a limited annotation budget.

1) FULLY SUPERVISED BASELINES

To isolate the effect of representation learning, we designed a fully supervised baseline to share the same encoder architecture as the contrastive model, while adding a task-specific decoder with skip connections, yielding a standard U-Net–like segmentation architecture. We trained a fully supervised segmentation model using only labeled data from scratch, without any form of self-supervised or unsupervised pretraining. To evaluate the performance under scarce annotation availability, we considered label budgets of 20% of the annotated training set (38 image-label pairs).

2) SELF-SUPERVISED LEARNING BASELINES

We include Model Genesis as a reference SSL baseline, that employs a set of pretext tasks based on image restoration and transformation prediction. For the SSL baselines, the encoder is first pretrained on the unlabeled data and subsequently fine-tuned for segmentation using the same labeled subsets as the fully supervised baseline.

3) FINE-TUNING AND BASELINE MODEL ARCHITECTURE AND TRAINING

All fine-tuning and baseline models were implemented in PyTorch [28] using the medical imaging framework MONAI [29]. The encoder–decoder backbone was based on the DynUNet architecture, configured for 3D volumetric processing with an input patch size of 128 × 128 × 32 voxels. The network employed four resolution levels with filter sizes of [32, 64, 128, 256], kernel size 3 × 3 × 3, and downsampling strides of (1, 2, 2, 2). The decoder used corresponding upsampling kernels of size two, and batch normalization was applied throughout. The network was trained to predict binary lesion segmentation masks (two output channels) without deep supervision. Fine-tuning was performed using combined Dice and cross-entropy loss (DiceCE) [30].

4) EVALUATION PROTOCOL

All methods were fine-tuned and evaluated using identical data splits, network architectures, and optimization settings. This protocol isolates the effect of the pretraining strategy and enables a direct comparison between fully supervised learning, existing SSL approaches, and the proposed anomaly-modulated spatial contrastive pretraining.

5) FINE-TUNING OPTIMIZATION SCHEDULE

Fine-tuning was conducted using a sawtooth learning rate schedule comprising three phases: an initial warm-up stage in which only the decoder parameters were optimized while the encoder remains frozen, a second warm-up phase in which all network parameters were jointly optimized, and a polynomial learning rate decay was applied for the remainder of the training. Linear warm-up was used in both initial phases, followed by polynomial decay with an exponent 0.9.

B. DATA AND EXPERIMENTAL SETUP

We conducted experiments on the MBHSeg dataset,[1] which consists of volumetric brain CT images with expert

[1] https://www.mbhseg.com/

annotations of microbleeds. The MBHSeg dataset is particularly suitable for this study because it provides both a large set of unlabeled volumes and a smaller annotated subset. This enabled the evaluation of representation learning under limited annotation budgets, which is the primary focus of this study. The dataset includes a mixture of healthy and pathological cases, allowing the anomaly estimation model to learn meaningful deviations from the normal anatomy, which is essential for the proposed anomaly-modulated pretraining framework. In this study, the task is formulated as a binary segmentation problem, distinguishing microbleed regions from the background. The dataset was partitioned into two subsets based on the availability of the annotations. A set of unlabeled images, including a total of 1443 healthy images and 563 images with pathology, were used to train the unpaired image translation model for anomaly map generation and contrastive pretraining of the encoder. To reduce the influence of noise and edge artifacts in the anomaly maps, lesion burden was estimated using only intermediate anomaly intensities [40, 90]. Input CT volumes were intensity-normalized using windowing with a range of [0, 100] Hounsfield units, followed by linear scaling to [0, 1] with clipping applied to out-of-range values. The annotated portion of the dataset contains 192 images. To assess robustness to label scarcity, fine-tuning with the same set of hyperparameters was repeated five times using randomly sampled 20% subsets (38 image-label pairs) of the labeled data, with performance averaged across folds. Each fold was split into training and validation at the subject level. The training set was used for supervised fine-tuning of the segmentation model, the validation set for model selection, and the test set for the final performance evaluation. All reported segmentation results were obtained on the held-out test set, which consisted of 40 images with non-zero ground truth masks.

The SEG-CQ500 [31] dataset was selected for the zero-shot cross-dataset generalization evaluation because it differs from the source dataset in the acquisition protocol, spatial resolution, and slice thickness. This makes it suitable for assessing the robustness of learned representations under domain shift. Unlike in-domain evaluation, zero-shot generalization testing provides insight into whether the pretrained features capture transferable anatomical and pathology-related information without requiring additional fine-tuning.

The implementation is publicly available at: https://github.com/ETRO-MIT/AMSA-BrainHemorrhageSegmentation.

VI. RESULTS

A. SEGMENTATION PERFORMANCE ON THE SOURCE DATASET

Table 1 summarizes the segmentation performances. AMSA outperforms fully supervised training and yields a modest but consistent improvement over spatial contrastive pretraining (Spatial CL). The performance of AMSA was comparable to that of reconstruction-based pretraining (Genesis), with no statistically significant difference observed.

TABLE 1. Segmentation performance on the source dataset. Dice scores are reported as mean ± standard deviation across folds. Statistical significance is computed with paired t-test and Wilcoxon signed-rank test compared to Spatial CL.

Method	Dice	p vs Spatial CL	p vs Genesis
Supervised	0.311 ± 0.273	–	–
Genesis	0.416 ± 0.241	–	–
Spatial CL	0.400 ± 0.264	–	–
AMSA (ours)	$\mathbf{0.417 \pm 0.249}$	0.022	0.08

Figure 3 illustrates the per-subject Dice distributions for all the evaluated training strategies. The proposed AMSA pretraining, implemented with $\beta = 2$ and $\sigma = 80$, after finetuning provides a statistically significant improvement over the fully supervised baseline trained from scratch, with an absolute improvement of 0.105. This improvement was statistically significant in the paired Wilcoxon signed-rank test ($p = 1 \times 10^{-5}$). The 95% bootstrap confidence interval for the mean Dice differences was [0.067, 0.144].

When compared to a contrastive pretraining baseline using spatial weighting only, Spatial CL, the proposed anomaly-modulated approach yielded a modest but consistent improvement. A one-sided paired Wilcoxon signed-rank test indicated statistical significance ($p = 0.022$), while the 95% bootstrap confidence interval for the mean Dice difference $[-0.004, 0.037]$ reflected substantial inter-subject variability, characteristic of small-lesion segmentation tasks, together with the modest average effect size.

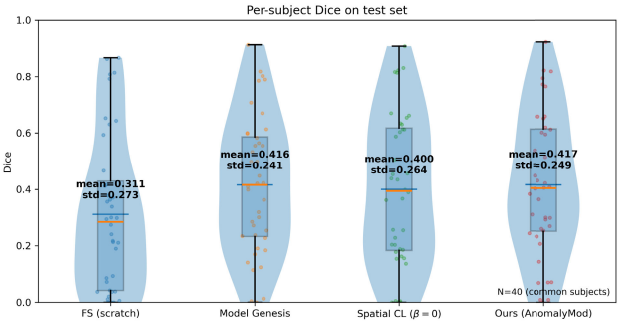


FIGURE 3. Per-subject Dice scores on the test set for four training strategies: fully supervised training from scratch (FS), Model Genesis pretraining, spatial contrastive learning without anomaly modulation ($\beta = 0$), and the proposed anomaly-modulated contrastive learning($\beta = 2$, $\sigma = 80$). Each point represents one subject ($N = 40$ subjects with non-empty ground-truth lesions). Violin plots depict the full Dice distribution, with embedded boxplots indicating median and interquartile range.

B. INTERACTION BETWEEN SPATIAL LOCALITY AND ANOMALY MODULATION

We further explored the influence of the spatial decay parameter σ and anomaly modulation strength β on segmentation performance, as shown in Figure 4. When σ is chosen to enforce local spatial interactions ($\sigma = 80$), increasing β

consistently improves Dice performance, with the best results observed for $\beta = 2$ and $\beta = 4$. In contrast, when σ is large ($\sigma = 200$), spatial weighting becomes nearly uniform and increasing β yields smaller gains or degrades performance. Furthermore, larger values of β under broad spatial weighting substantially increased performance variability, as reflected by the higher standard deviation in Dice. These results empirically validated our design principles, demonstrating that anomaly modulation enhances representation learning only when coupled with a sufficiently local spatial prior.

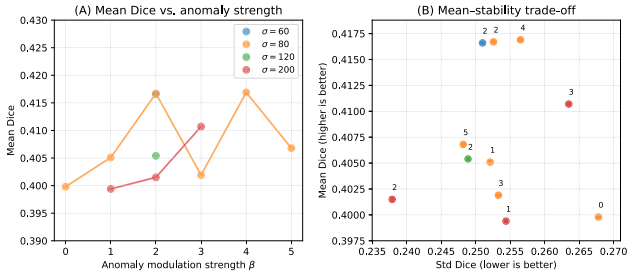


FIGURE 4. Effect of anomaly modulation strength β and spatial decay σ on segmentation performance. (A) Mean Dice score as a function of β, shown separately for each σ. (B) Mean–stability trade-off, where each point corresponds to a (β, σ) configuration (higher is better on the y-axis; lower is better on the x-axis). The number next to each dot represents a β value. Moderate anomaly modulation combined with a local spatial prior ($\sigma = 80$) yields the best balance between accuracy and robustness, supporting the proposed loss design principles. Performance improves when anomaly modulation is coupled with a local spatial prior ($\sigma = 80$), while broader spatial weighting ($\sigma = 200$) reduces gains and increases variability.

The ablation results in Figure 4 highlight the necessity of coupling anomaly modulation with the spatial locality. When the spatial decay parameter σ is matched to the inter-location spacing of the patch grid, contrastive interactions remain anatomically meaningful, allowing anomaly-derived signals to encode lesion-related information without disrupting spatial structure. Conversely, when σ is large, spatial interactions become effectively global, and increasing the anomaly modulation strength β amplifies the noise in the anomaly estimates, leading to unstable training and reduced performance.

C. STRATIFIED ANALYSIS BY LESION BURDEN ON THE SOURCE DATASET

To further analyze model behavior, we evaluated per-subject changes in Dice score between AMSA and spatial contrastive pretraining across different lesion burden groups as shown in Figure 5. This analysis assesses whether anomaly modulation provides differential benefits that are not captured by the aggregate metrics.

The results indicated that improvements were most pronounced in the low-lesion regime. In the small-lesion group (<5 mL), AMSA achieved a mean Dice gain of approximately +0.025 compared with spatial-only contrastive learning, although this trend did not reach statistical significance after multiple-comparison correction. For medium (5)–20 mL) and high (>20 mL) lesion burdens, performance remained stable, with smaller yet stable gains (mean Δ Dice +0.012), suggesting a tendency toward improved performance in low-lesion cases, although this effect was not statistically significant, while preserving performance for larger lesions. No monotonic relationship between lesion burden and performance change is observed (Spearman $\rho = -0.11$, $p = 0.50$), indicating that anomaly modulation does not bias representations toward severe pathology. Overall, these findings suggest that anomaly modulation preferentially benefits cases with a small lesion burden, where segmentation is more challenging, while preserving the performance in moderate and high-burden cases. Although overall performance between AMSA and Genesis is comparable, stratified analysis indicates a tendency to improve the performance of AMSA in low-burden cases, where lesions are small and more challenging to detect. In this subgroup, AMSA showed a positive mean Dice difference (approximately +0.03); although statistical significance was not reached due to the limited sample size.

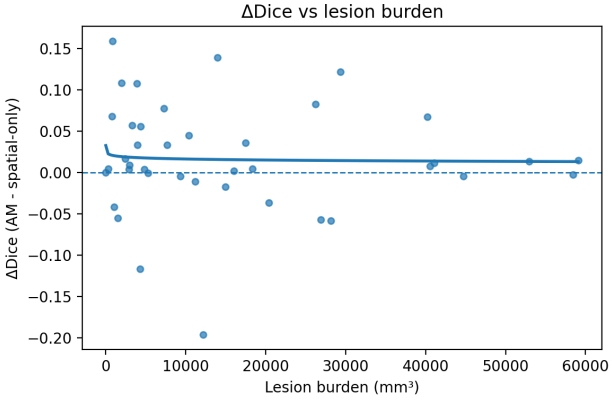


FIGURE 5. Per-subject Dice difference between AMSA and Spatial CL stratified by ground-truth lesion burden. Each point represents a single subject, grouped by lesion volume: small (<5 mL), medium (5–20 mL), and large (>20 mL). Positive values indicate improved performance of AMSA.

D. EMBEDDING SPACE ANALYSIS

To better understand the effect of anomaly-modulated contrastive pretraining on the learned representation space, we conducted qualitative and quantitative analyses of the encoder embeddings. For each model configuration, feature vectors were extracted from patches in the held-out test set and projected to two dimensions using the t-SNE. Figure 6 illustrates a clear and well-structured organization of the embedding space: patches originating from the same anatomical location form compact and distinct clusters, confirming that the spatial component of the loss successfully enforces anatomical coherence. Figure 7 illustrates that while embeddings remain primarily organized by spatial location, lesion-related information manifests as consistent shifts within location-specific clusters rather than forming independent clusters, suggesting that anomaly modulation

introduces lesion-related information without disrupting anatomical coherence.

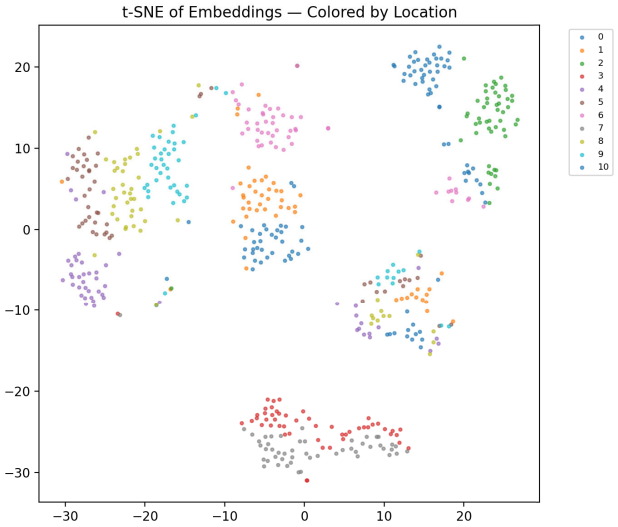


FIGURE 6. t-SNE of embeddings colored by patch location label. The colors represent 11 different location labels.

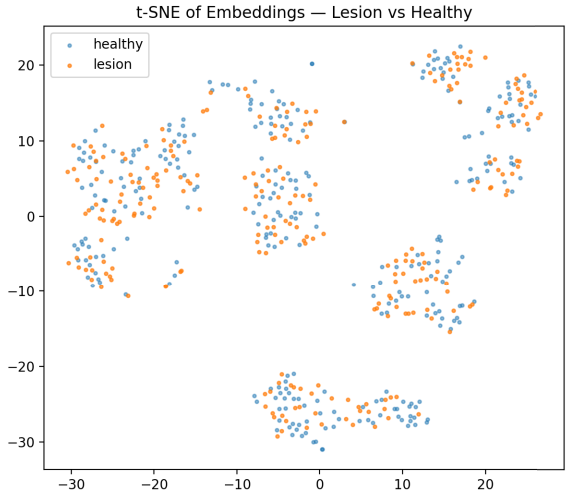


FIGURE 7. t-SNE of embeddings colored by pathology label of the patch. The colors represent pathology labels ("healthy" vs "lesion").

To complement the qualitative visualization, we also performed linear probe analysis where two simple linear logistic regression classifiers were trained on frozen embeddings of the model Genesis and of AMSA. AMSA embeddings achieved an AUC of 0.72 for lesion versus healthy patch classification compared to 0.64 for Model Genesis, while preserving strong anatomical organization (location accuracy 0.96 for both methods), Figure 8. Linear probe analysis revealed that anomaly-modulated contrastive pretraining produces representations where lesion information is more linearly accessible. These results suggest that the anomaly modulation enhances pathology-related information without compromising anatomical consistency. Although lesion and healthy patches are not cleanly separable in the low-dimensional t-SNE projection for the AMSA model, the consistent within-cluster structure supports the observed linear probe performance, indicating that pathology-related information is encoded in the embedding space in a location-aware manner. The proposed framework structures the representation space such that the anatomy is preserved while the pathology is made more salient.

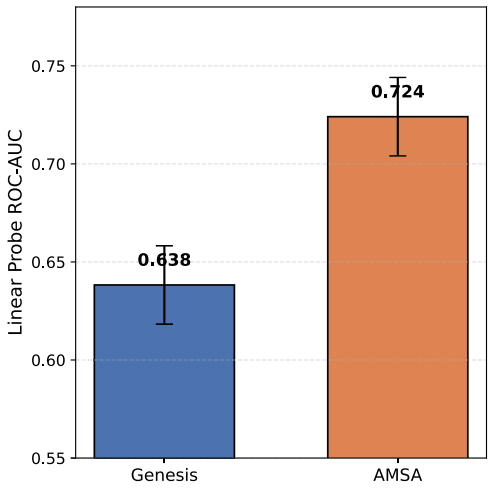


FIGURE 8. Linear probe evaluation of representation quality. A logistic regression classifier trained on frozen embeddings shows improved lesion discrimination for AMSA (AUC = 0.72) compared to Model Genesis (AUC = 0.64).

E. LESION DETECTION EVALUATION

While segmentation evaluates voxel-level accuracy, lesion-wise detection provides complementary insight into the ability of the model to identify clinically relevant lesion candidates. To evaluate lesion-level detection performance, voxel-wise probability maps produced by each model were converted into discrete lesion candidates. Specifically, for each test volume, the predicted probability maps were thresholded at multiple operating points to obtain the binary masks. Connected component analysis was then applied to identify individual lesion candidates, and small components below a predefined minimum volume were discarded to suppress noise-induced detections. Each remaining connected component was assigned a confidence score corresponding to the maximum predicted probability within that component. A predicted lesion was considered a true positive if it overlapped with any ground-truth lesion mask; otherwise, it was counted as a false positive. Ground-truth lesions that were not matched by any prediction were counted as false negatives. This procedure enabled a lesion-wise evaluation independent of voxel-level segmentation metrics. Free-response receiver operating characteristic (FROC) curves were generated by varying the decision threshold applied to lesion confidence scores. FROC analysis provides a

complementary perspective to the Dice score by evaluating lesion detection performance at different false-positive rates. This is particularly important in clinical settings, where the ability to detect lesions with a low number of false positives is critical. In this context, higher recall in the low false-positive regime indicates that the model is more effective at prioritizing clinically relevant lesion candidates. For each threshold, lesion recall was computed as the ratio of correctly detected lesions to the total number of ground-truth lesions, while the number of false positives per scan was calculated by averaging the number of unmatched predicted lesions across all test volumes. The resulting FROC curves characterize the trade-off between sensitivity and false-positive burden across different operating points, as shown in Figure 9. On the source dataset, AMSA achieves higher recall than both Genesis and Spatial CL at fewer than five false positives per scan, highlighting lesion-level detection differences that are not fully reflected by Dice score alone.

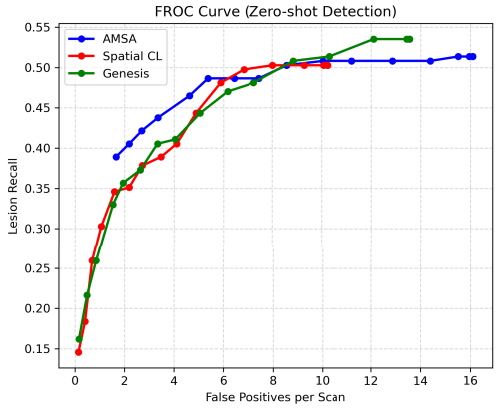


FIGURE 9. FROC curves comparing lesion detection performance of AMSA, Spatial CL (AMSA without anomaly modulation), and Genesis on the source dataset. AMSA achieves higher recall in the low false-positive regime, indicating improved ranking of high-confidence lesion candidates, while all methods converge at higher false-positive rates.

F. ZERO-SHOT CROSS-DATASET GENERALIZATION

The purpose of this experiment was to analyze how the learned representations generalize to unseen data under a domain shift. In particular, this setting allowed us to examine differences in model behavior, such as sensitivity to lesion presence and robustness to distribution changes, without the confounding effect of domain-specific fine-tuning.

We evaluated the generalization capability of the pretrained models in a zero-shot setting on the SEG-CQ500 dataset. As a test set we used 32 3D CT images from the SEG-CQ500 dataset, 512×512 in-plane resolution and slice thickness varying from 0.5 to 5 mm with corresponding ground truth masks. Test images were affinely registered to the template used during pretraining. Without any additional fine-tuning, the models trained on the source dataset were directly applied to the target dataset, which differs in acquisition protocol and slice thickness. All volumes were preprocessed according to the original AMSA preprocessing pipeline. Inference was

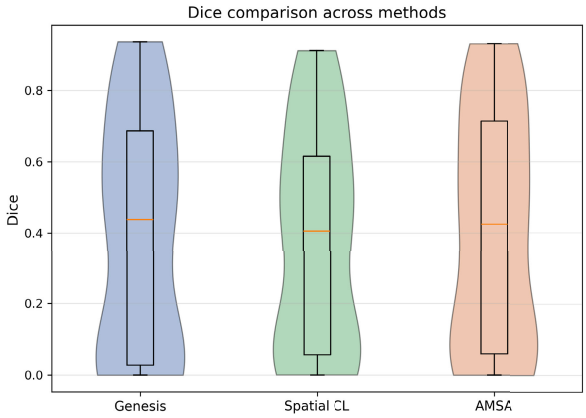


FIGURE 10. Comparison of zero-shot results on Seg-CQ500 dataset from 3 models: Genesis, Spatial CL, AMSA. Violin plots depict the full Dice distribution, boxplots indicate median and interquartile range.

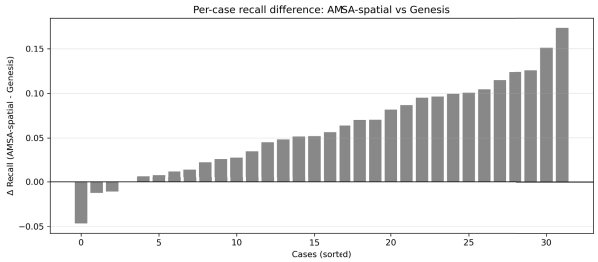


FIGURE 11. Per-case difference in recall (sensitivity) between Spatial CL and Genesis models on the SEG-CQ500 dataset. Each bar represents a single case, sorted by increasing difference. Positive values indicate higher recall for Spatial CL, while negative values indicate better performance of Genesis. The results show that Spatial CL achieves higher recall in the majority of cases, suggesting improved lesion detection capability under domain shift.

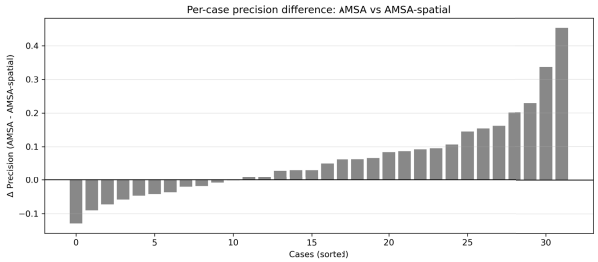


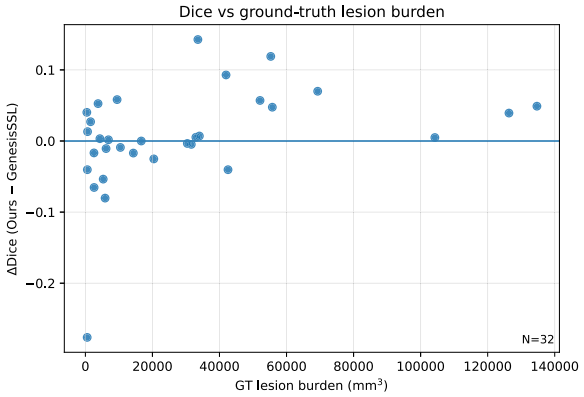
FIGURE 12. Per-case difference in precision between AMSA and Spatial CL models on the SEG-CQ500 dataset. Each bar corresponds to a case, sorted by increasing difference. Positive values indicate higher precision for AMSA, whereas negative values indicate improved precision for Spatial CL. AMSA partially restores precision while maintaining competitive recall.

performed using a sliding-window strategy with a patch size of (128 × 128 × 32). For inference we used five models for each setup (AMSA, Spatial CL and Genesis), obtained via fine-tuning on 5 folds in the previous experiments on the source dataset.

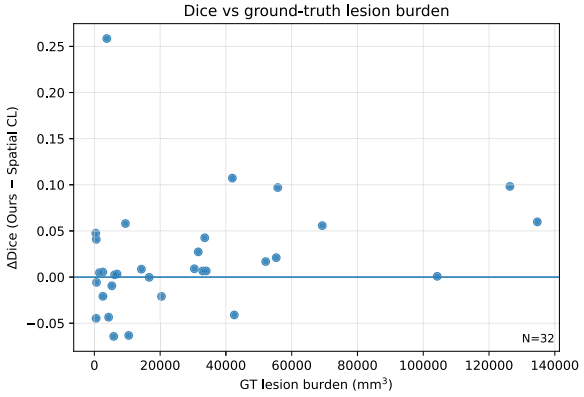
Performance is evaluated using Dice similarity coefficient as illustrated in Figure 10, and precision, and sensitivity (recall) in Table 2. The metrics were computed on a per-case basis and averaged across five folds. For comparative analysis, probability thresholds were swept independently

TABLE 2. Zero-shot cross-dataset segmentation performance comparison ($N = 32$) for Seg-CQ500 dataset. Best values are in bold. Δ indicates mean difference relative to comparison method. *p*-values are from paired Wilcoxon signed-rank tests.

Metric	Genesis	AMSA-spatial	AMSA	Δ AMSA vs Genesis	Δ AMSA vs Spatial
Dice	0.3867 ± 0.3302	0.3698 ± 0.3052	$\mathbf{0.4015 \pm 0.3239}$	$+0.0147$ ($p = 0.164$)	$\mathbf{+0.0317}$ ($p = 0.0069$)
Precision	$\mathbf{0.5203 \pm 0.3616}$	0.4184 ± 0.3072	0.4801 ± 0.3427	-0.0402 ($p = 0.0037$)	$\mathbf{+0.0616}$ ($p = 0.0083$)
Recall	0.3600 ± 0.3323	$\mathbf{0.4192 \pm 0.3431}$	0.4156 ± 0.3443	$\mathbf{+0.0556}$ ($p = 0.0001$)	-0.0036 ($p = 0.5988$)



(a) The per-subject change in Dice score between anomaly-modulated spatial contrastive pretraining and model Genesis.



(b) The per-subject change in Dice score between anomaly-modulated spatial contrastive pretraining and a spatial-only contrastive baseline.

FIGURE 13. Lesion burden stratified analysis of zero-shot cross-dataset (SEG-CQ500 dataset) segmentation results reveals that performance gains are mostly observed in subjects with larger lesion burden, where AMSA pretraining achieves improvements over both reconstruction-based and spatial-only contrastive approaches.

for each model, and the operating point corresponding to the highest mean Dice score on the target dataset was used for reporting lesion-level sensitivity and false-positive characteristics.

The zero-shot cross-dataset evaluation revealed that AMSA improved Dice over its spatial-only variant by $+0.0317$ ($p = 0.0069$), primarily by recovering precision while maintaining comparable recall. Both AMSA and the spatial-only model increased recall over Genesis ($+0.056$, $p < 0.001$), indicating enhanced sensitivity to lesion presence, Figure 11. However, this gain was accompanied by a substantial reduction in precision for the spatial-only model ($\Delta -0.1019$, $p = 0.0004$), reflecting over-segmentation. Anomaly-aware modulation partially restored precision ($+0.0616$ vs spatial-only, $p = 0.0083$), resulting in a more balanced trade-off between precision and recall compared to the spatial-only variant, while remaining competitive with Genesis, Figure 12.

Lesion-burden stratified analysis showed that improvements were concentrated in cases with larger lesion volumes as demonstrated in Figure 13, consistent with a positive correlation between lesion burden and performance gain. Spatial pretraining primarily enhances the detection of prominent structures, whereas anomaly-aware objectives refine these representations by suppressing false positives, particularly in high-burden regimes. In contrast, small-lesion cases remain challenging.

VII. DISCUSSION

In this study, we introduced an anomaly-modulated spatial contrastive learning framework for self-supervised lesion segmentation. In the limited label setting, AMSA improves over fully supervised training and spatial-only contrastive pretraining, while achieving a performance comparable to reconstruction-based pretraining with Model Genesis. Embedding analysis and linear probing further suggest that anomaly modulation increases the accessibility of pathology-related information while preserving the anatomical organization. Lesion-level evaluation provides additional insight beyond the voxel-level Dice score. On the source dataset, AMSA achieves higher recall in the low false-positive regime than both Genesis and Spatial CL, indicating improved ranking of clinically relevant lesion candidates in the low false-positive regime. Although the absolute Dice improvements are modest, such behavior is expected in lesion segmentation tasks involving small and sparse targets, where small numerical gains may correspond to clinically meaningful improvements in lesion detection. Prior clinical studies suggest that hemorrhagic lesions below approximately 5–10 mL may be missed in 10–30% of cases [32], [33]. Zero-shot cross-dataset evaluation reveals a distinct operating behavior under domain shift. Compared with Genesis, AMSA achieves higher recall but lower precision, indicating increased sensitivity to lesion presence at the cost of additional false positives. In contrast, Spatial CL produces the strongest recall increase but also exhibits substantial over-segmentation. AMSA partially restores precision relative to Spatial CL while maintaining comparable recall, resulting in a more balanced precision–recall trade-off. One possible explanation is that anomaly-aware modulation constrains overly broad similarity matching between anatomically similar regions, thereby reducing over-segmentation tendencies observed in purely spatial contrastive learning. Reconstruction-based SSL methods

such as Model Genesis remain strong baselines for lesion segmentation [34] and achieved comparable performance in our experiments. However, the proposed anomaly modulation consistently improves upon purely spatial contrastive learning and substantially narrows the gap between contrastive and reconstruction-based SSL. These findings suggest that incorporating weak pathology-aware signals into contrastive pretraining may provide a complementary alternative to reconstruction objectives for lesion-sensitive representation learning.

This study has several limitations. First, anomaly maps are derived from a generative model and therefore provide only a noisy proxy for pathology. Second, we intentionally used a lightweight encoder architecture to isolate the effect of the proposed loss, and stronger backbones may further improve performance. Finally, although zero-shot evaluation provides initial evidence of generalization, broader validation across datasets, modalities, and lesion types is necessary. Future work will explore improved anomaly estimation, stronger backbone architectures, and hybrid approaches combining anomaly-modulated contrastive learning with reconstruction-based objectives.

REFERENCES

[1] K. C. Teo, C. W. H. Chan, and A. H. Y. Lee, "Location-specific hematoma volume cutoff and clinical outcomes in intracerebral hemorrhage," *Stroke*, vol. 54, no. 7, pp. 2056–2065, Jul. 2023.

[2] R. Meyrat, E. Vivian, B. Dulaney, and J. C. B. Gutierrez, "Enhanced prediction of in-hospital mortality in intracerebral hemorrhage: Impact of serial neurological and radiological reassessment with the ICH Score at 6 hours postadmission," *J. Neurosurg.*, pp. 1–10, Oct. 2025.

[3] J.-Y. Lee, C. King, D. Stradling, M. Warren, D. Nguyen, J. Lee, M. A. Riola, R. Montoya, D. Patel, V. H. Le, S. J. Welbourne, and S. C. Cramer, "Influence of hematoma location on acute mortality after intracerebral hemorrhage," *J. Neuroimaging*, vol. 24, no. 2, pp. 131–136, Mar. 2014.

[4] A. Morotti, Q. Li, J. Nawabi, F. Mazzacane, F. Schlunk, A. Shoamanesh, G. Busto, A. Cavallini, F. Palmerini, M. Paciaroni, E. M. Gurol, A. Viswanathan, I. Casetta, L. Piccolo, E. Fainardi, S. M. Greenberg, A. Padovani, A. Zini, J. Rosand, J. P. Broderick, D. Dowlatshahi, and J. N. Goldstein, "Volume tolerance and prognostic impact of hematoma expansion in deep and lobar intracerebral hemorrhage," *Stroke*, vol. 56, no. 5, pp. 1224–1231, May 2025.

[5] P. Panni, F. Simionato, R. Cao, A. Pedicelli, E. Marchese, A. Caricato, A. Alexandre, A. Feletti, M. Testa, P. Zanatta, N. Gitti, S. Piva, D. Mardighian, V. Semeraro, G. Nardin, E. Lozupone, G. Paiano, E. Picetti, V. Montanaro, M. Petranca, C. Bortolotti, A. Scibilia, L. Cirillo, R. Aspide, A. L. Lanterna, A. Ambrosi, P. Mortini, M. L. Azzolini, M. R. Calvi, and A. Falini, "Hemorrhage volume drives early brain injury and outcome in poor-grade aneurysmal SAH," *AJNR Am. J. Neuroradiol.*, vol. 45, no. 4, pp. 393–399, Apr. 2024.

[6] W.-C. Wang, E. Ahn, D. Feng, and J. Kim, "A review of predictive and contrastive self-supervised learning for medical images," *Int. J. Autom. Comput.*, vol. 20, no. 2, pp. 1–18, Apr. 2023.

[7] C. Zhang, H. Zheng, and Y. Gu, "Dive into the details of self-supervised learning for medical image analysis," *Med. Image Anal.*, vol. 87, Jun. 2023, Art. no. 102807.

[8] C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised visual representation learning by context prediction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1422–1430.

[9] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.

[10] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, 2018.

[11] Z. Zhou, V. Sodha, M. M. R. Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, and J. Liang, "Models genesis: Generic autodidactic models for 3D medical image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 11767, 2019, pp. 384–393.

[12] K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu, "Contrastive learning of global and local features for medical image segmentation," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 33, 2020, pp. 12546–12558.

[13] L. Wu, J.-X. Zhuang, and H. Chen, "VoCo: A simple-yet-effective volume contrastive learning framework for 3D medical image analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, 2024, pp. 22873–22882.

[14] V. Agarwal, Y. Tang, J. Xiao, and R. M. Summers, "Weakly supervised lesion co-segmentation on CT scans," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Iowa City, IA, USA, 2020, pp. 203–206.

[15] X. Zhang, J. Wang, J. Wei, X. Yuan, and M. Wu, "A review of non-fully supervised deep learning for medical image segmentation," *Information*, vol. 16, no. 6, p. 433, May 2025.

[16] J. Li, Y. Zheng, K. Wu, J. Shi, F. Xie, and Z. Jiang, "Lesion-aware contrastive representation learning for histopathology whole slide images analysis," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022* (Lecture Notes in Computer Science), vol. 13435, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds., Cham, Switzerland: Springer, 2022, pp. 292–302.

[17] Y. Han, C. Chen, A. Tewfik, B. Glicksberg, Y. Ding, Y. Peng, and Z. Wang, "Knowledge-augmented contrastive learning for abnormality classification and localization in chest X-rays with radiomics using a feedback loop," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2022, pp. 1789–1798.

[18] D. Zimmerer, S. A. A. Kohl, J. Petersen, F. Isensee, and K. H. Maier-Hein, "Context-encoding variational autoencoder for unsupervised anomaly detection," in *Proc. Med. Imag. Deep Learn. (MIDL)*, vol. 102, 2019.

[19] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.

[20] Y. Li, Q. Lao, Q. Kang, Z. Jiang, S. Du, S. Zhang, and K. Li, "Self-supervised anomaly detection, staging and segmentation for retinal images," *Med. Image Anal.*, vol. 87, Jul. 2023, Art. no. 102805.

[21] N. Pinon and C. Lartizien, "OCSVM-guided representation learning for unsupervised anomaly detection," 2025, *arXiv:2507.21164*.

[22] Y. Song and J. Cheng, "Self-supervised anomaly detection of medical images based on dual-module discrepancy," in *Proc. ACM Multimedia Asia*, Dec. 2023, pp. 1–7.

[23] S. Hansen, S. Gautam, R. Jenssen, and M. Kampffmeyer, "Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels," *Med. Image Anal.*, vol. 78, May 2022, Art. no. 102385.

[24] Z. Sun, Y. Gu, Y. Liu, Z. Zhang, Z. Zhao, and Y. Xu, "Position-guided prompt learning for anomaly detection in chest X-rays," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2024, pp. 567–577.

[25] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 319–345.

[26] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[27] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn. (ICML)*, vol. 119, 2020, pp. 1597–1607.

[28] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, 2019, pp. 8026–8037.

[29] M. J. Cardoso et al., "MONAI: An open-source framework for deep learning in healthcare," Nov. 2022, *arXiv:2211.02701*.

[30] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[31] A. Spahr, J. Ståhle, C. Wang, and M. Kaijser, "Label-efficient deep semantic segmentation of intracranial hemorrhages in CT-scans," *Front. Neuroimaging*, vol. 2, Jul. 2023, Art. no. 1157565.

[32] J. M. Wardlaw, S. L. Keir, J. Seymour, S. Lewis, and P. A. Sandercock, "Neuroimaging in acute stroke," *Lancet Neurol.*, vol. 3, no. 1, pp. 27–41, 2004.

[33] P. Reschke, L. D. Gruenewald, V. Koch, J. Gotta, E. Höhne, A.-I. Nica, I. Yel, T. D'Angelo, G. Ascenti, G. M. Bucolo, L. R. M. Lanzafame, E. Jubica, K. Eichler, T. J. Vogl, and C. Booz, "AI assistance improves radiology resident reader performance in CT diagnosis of intracranial hemorrhage," *La radiologia medica*, vol. 131, no. 2, pp. 280–291, Oct. 2025.
[34] T. Wald, C. Ulrich, J. Suprijadi, S. Ziegler, M. Nohel, R. Peretzke, G. Köhler, and K. Maier-Hein, "An OpenMind for 3D medical vision self-supervised learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2025, pp. 23839–23879.

MARYNA KVASNYTSIA received the M.S. degree in medical radiophysics from the Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, in 2006. She is currently pursuing the joint Ph.D. degree with the Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Brussels, and IMEC Belgium, focusing on self-supervised learning for medical image segmentation. She has over 15 years of experience in the applied research and development of biomedical image analysis pipelines. From 2007 to 2015, she was with Materialise NV, Belgium, where she conducted applied research on medical image analysis, including cardiovascular CT quantification, micro-CT analysis, and biomedical software development. She held research positions with KU Leuven, Icometrix, and ETRO–VUB–IMEC. Her research interests include self-supervised learning for medical image analysis and machine learning for biomedical applications.

ABEL DÍAZ BERENGUER received the B.Sc. (Eng.) degree in informatics sciences and the M.Sc. (Eng.) degree in applied informatics from the University of Informatics Sciences (UCI), Havana, Cuba, in 2009 and 2014, respectively, and the Ph.D. degree in engineering from Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 2021. He is a Postdoctoral Research Engineer with the Department of Electronics and Informatics (ETRO), VUB. His current research interests include computer vision, representation learning, machine learning, multimodal foundation models, and agentic AI systems, particularly for healthcare applications, such as medical imaging, clinical decision support, personalized medicine, and biomedical discovery.

HICHEM SAHLI is a Professor of computer vision and machine learning with the Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB); and a Principal Scientist with the Interuniversity Microelectronics Centre (IMEC), Belgium. His contributions span affective computing, multimodal signal analysis, and health informatics, with work on medical image analysis, disease progression modeling, diagnosis and treatment outcome prediction, and electronic health record coding. He has authored over 190 journal publications and completed 32 Ph.D.s, since 2000. His research interests include the theory and development of machine learning and signal processing algorithms for computer vision, including radar, image, and video analysis.

JEF VANDEMEULEBROUCKE received the master's degree in electronic engineering from the University of Ghent, Belgium. In Granada, Spain, he specialized in artificial intelligence for one year. He performed postgraduate training in numerical optimization techniques at the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil. His doctoral research focused on lung motion estimation and modeling for image-guided radiation therapy, conducted in collaboration with the Creatis Laboratory, University of Lyon 1, France, and the Center for Machine Perception, Czech Technical University, Prague, Czech Republic. He is currently an Associate Professor of medical image analysis with the Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Belgium. He is also an Affiliated Researcher with IMEC, an international research and innovation hub in nanoelectronics and digital technologies. His current research interests include medical image analysis for applications in computer-aided diagnosis and image-guided interventions, with a focus on thoracic, whole-body, and dynamic imaging for oncology and musculoskeletal pathologies.