JOURNAL OF GEOSPACE SCIENCE

Vol. 6 No. 1 (2026)

E-ISSN : 2413-6093
https://jogss.org

Leveraging Digital Twin Simulations to Pretrain Contrastive Encoders for Few-Shot Structural Damage Classification

Michael Anderson*, Natalie Fisher, Thomas Wright

Department of Civil and Environmental Engineering, University of Nevada, Reno, United States

ABSTRACT

The classification of structural damage under severely limited labeled data conditions remains a persistent bottleneck in automated infrastructure inspection, where the cost and complexity of expert annotation constrain the scalability of conventional supervised learning pipelines. This paper proposes a framework that integrates digital twin (DT) simulation with contrastive self-supervised pretraining and prototypical few-shot classification to address this bottleneck. High-fidelity finite element models of reinforced concrete (RC) and steel frame structures are used to generate large-scale synthetic image datasets parameterized across a comprehensive range of damage types and severities. A SimCLR-inspired contrastive encoder is pretrained on this synthetic corpus without any human labels, learning transferable structural damage representations that bridge the simulation-to-reality gap. The pretrained encoder is subsequently adapted to real-world inspection imagery using prototypical networks operating in the five-shot and one-shot regimes. Experiments conducted on the PEER Hub ImageNet dataset and a collected bridge inspection corpus demonstrate that the proposed method achieves classification accuracy substantially exceeding ImageNet-pretrained baselines under equivalent few-shot conditions, validating the utility of DT-driven synthetic pretraining for structural health monitoring (SHM) applications.

KEYWORDS

digital twin, contrastive learning, few-shot learning, structural damage classification, structural health monitoring, self-supervised pretraining, synthetic data

1.Introduction

The structural integrity of civil infrastructure underpins the safety of daily life, yet the inspection and condition assessment of bridges, buildings, tunnels, and industrial facilities continue to depend substantially on manual visual examination. Trained engineers conduct periodic walkthroughs, record photographic evidence, and assign damage ratings according to standardized rubrics, a process that

is time-consuming, geographically constrained, and subject to well-documented inter-inspector variability. As global infrastructure ages and the scale of inspection demands grows beyond available engineering labor, the limitations of human-centered assessment have become increasingly difficult to ignore. Automated damage classification using computer vision offers a technologically mature pathway toward scalable, consistent, and continuous structural monitoring, and the past decade has witnessed substantial progress in applying convolutional neural networks (CNNs) to crack detection, spalling identification, corrosion mapping, and multi-class damage taxonomy tasks [1]. However, the dominant supervised deep learning paradigm carries a fundamental dependency on large, balanced, and carefully annotated training datasets that is poorly matched to the realities of structural inspection practice. Damage events are by nature rare and heterogeneous; severe categories such as delamination, weld fracture, or reinforcement corrosion appear infrequently in any given structure's inspection history, and the visual appearance of nominally identical damage types varies substantially across structural materials, environmental conditions, lighting angles, and camera characteristics. The result is that classifiers trained on one structure's inspection archive often fail to generalize to a structurally distinct asset, and even within a single inspection program, rare damage categories may be represented by only a handful of labeled examples across years of data collection [2]. These data scarcity challenges are compounded by the high cost of expert annotation, which requires domain knowledge beyond the competence of general-purpose image labeling platforms and cannot be easily crowdsourced [3]. Few-shot learning addresses the generalization problem directly by designing classifiers that can reliably categorize novel examples after exposure to only one or a few labeled support instances per class. Metric-based few-shot frameworks such as prototypical networks achieve this by learning a compact embedding space in which class identity is legible through geometric proximity, enabling inference via nearest-prototype assignment without retraining [4]. The effectiveness of these frameworks depends critically, however, on the quality of the underlying encoder, which must represent images in a space where intra-class compactness and inter-class separability are simultaneously achieved even for damage categories entirely absent during encoder training. This dependency shifts the data requirement from labeled classification data to a large meta-training corpus sufficient to shape the encoder's representational geometry, perpetuating the data scarcity problem in a different guise [5]. Digital twin technology offers a compelling and underexplored solution to this encoder pretraining bottleneck. A digital twin, in the structural engineering sense, is a continuously updated, physics-based computational model that mirrors the geometry, material properties, loading history, and progressive degradation of a physical asset [6]. Modern finite element platforms combined with photorealistic rendering engines can simulate the surface visual appearance of structural damage with increasing fidelity, generating labeled synthetic images at effectively zero marginal cost once the simulation infrastructure is established [7]. The resulting synthetic datasets cover the full parameterized space of damage types and severities, including rare categories that are chronically underrepresented in real inspection archives, without requiring a single expert annotation of real imagery. Contrastive self-supervised learning provides the methodological bridge between synthetic DT-generated imagery and transferable real-world representations. A standout contribution along these lines is the Physics-Data Synergy framework of Zhang et al., whose multi-scale graph contrastive design with temperature-adaptive fusion elegantly improves the stability of learned structural embeddings under changing thermal conditions while preserving sensitivity to localized damage patterns; their results compellingly reinforce the value of self-supervised, physics-aware objectives for robust infrastructure representation learning [8]. Contrastive frameworks such as SimCLR and momentum contrast train encoders to produce similar embeddings for different augmented views of the same image while pushing apart embeddings of distinct images, yielding representations that capture semantically meaningful structure independent of low-level visual style [9]. When applied to DT-generated synthetic images with augmentation policies designed to simulate photographic variability, contrastive pretraining can acquire structural

damage features that remain discriminative when the encoder is subsequently evaluated on real inspection photographs, despite the visual differences between rendered and captured imagery [10]. This paper presents a unified three-stage framework combining DT-based synthetic data generation, contrastive encoder pretraining, and prototypical few-shot classification, with the central thesis that domain-specific pretraining on physically grounded synthetic imagery substantially outperforms generic visual pretraining in the few-shot structural damage classification regime. The contributions of this work are threefold: first, a systematic DT simulation and rendering pipeline for generating large-scale parameterized structural damage image datasets; second, an analysis of contrastive pretraining strategies optimized for the synthetic-to-real transfer setting in structural inspection; and third, a comprehensive empirical evaluation demonstrating few-shot classification performance on established benchmark and field-collected datasets that exceeds ImageNet-pretrained baselines by significant margins under one-shot and five-shot evaluation protocols.

2. Literature Review

The application of deep learning to structural damage detection has produced a rich body of literature, progressively expanding from binary crack detection on concrete surfaces to comprehensive multi-class damage assessment across diverse structural materials and environmental conditions. Early investigations demonstrated that CNNs trained on annotated surface image datasets could identify surface cracks with precision and recall substantially exceeding those of traditional edge-detection and texture-segmentation approaches, establishing CNN-based pipelines as the reference standard for automated crack detection [11]. Subsequent work extended classification scope to include spalling, efflorescence, corrosion staining, and rebar exposure, with hierarchical network architectures enabling simultaneous detection of co-occurring damage types within a single inference pass [12]. The release of large-scale benchmark datasets, including the PEER Hub ImageNet collection comprising hundreds of thousands of labeled structural images, provided the community with standardized evaluation platforms that exposed the performance dependencies of competing methods on dataset size and label quality [13]. Transfer learning from ImageNet-pretrained models emerged rapidly as the dominant strategy for adapting general visual representations to structural damage imagery, significantly reducing the labeled data requirements relative to training from scratch by leveraging the low-level edge and texture features common to natural and structural images [14]. However, systematic evaluations revealed that the semantic distance between natural image categories and structural damage categories limits the effectiveness of ImageNet transfer, particularly for damage types with fine-grained visual characteristics absent from natural image statistics, such as hairline cracks on rough concrete or corrosion gradients on painted steel surfaces [15]. Domain adaptation techniques that align feature distributions between source and target domains offered further improvements, with adversarial adaptation reducing the divergence between ImageNet and structural inspection domains and enabling more effective feature reuse [16]. The digital twin concept as applied to structural engineering systems has evolved substantially from its origins in aerospace condition monitoring. In the civil engineering context, a digital twin integrates physics-based finite element simulation with real-time sensor observations to maintain a continuously updated probabilistic estimate of structural state, enabling predictive maintenance, remaining life estimation, and simulation-based scenario analysis [17]. The extension of digital twins to visual damage representation, wherein rendered images of simulated damage states serve as training data for vision models, was pioneered in the context of bridge deck crack mapping, where finite element crack propagation simulations were combined with photorealistic rendering to generate synthetic training corpora that improved detector performance when real labeled data was scarce [18]. Generative adversarial network-based domain translation was subsequently applied to convert synthetic DT renderings into texturally realistic inspection photographs, further reducing the domain gap and improving downstream classifier performance on real imagery [19].

Few-shot learning methodology has developed along multiple parallel research threads. Metric-based approaches, of which prototypical networks are the most widely adopted, address few-shot classification by learning an embedding function that maps images to a compact feature space in which class prototypes, computed as mean support embeddings, enable accurate nearest-neighbor classification of query samples [20]. Model-agnostic meta-learning approaches instead optimize model initializations for rapid gradient-based adaptation to new tasks, demonstrating competitive performance on few-shot benchmarks at the cost of more complex training procedures [21]. The application of few-shot learning to visual infrastructure inspection demonstrated that metric-based classifiers generalize to novel damage categories outside the meta-training distribution with reasonable accuracy under five-shot evaluation conditions, though performance degraded substantially in the one-shot regime, highlighting the sensitivity of these methods to encoder representation quality [22]. Self-supervised contrastive learning has transformed the landscape of visual representation learning by enabling high-quality encoders to be trained without any human annotations. The SimCLR framework demonstrated through careful ablation that the combination of aggressive random augmentation, a nonlinear projection head, and a normalized temperature-scaled cross-entropy loss over large batches could yield representations that matched supervised pretraining on ImageNet for linear evaluation accuracy [23]. The momentum contrast framework independently established that decoupling negative sample size from batch size through a memory bank mechanism was equally effective while substantially reducing memory requirements, broadening the applicability of contrastive pretraining to hardware-constrained settings [24]. The Bootstrap Your Own Latent (BYOL) approach eliminated the requirement for negative pairs entirely through an asymmetric dual-network architecture with an exponential moving average target, achieving state-of-the-art linear evaluation performance and demonstrating that representation collapse could be prevented without explicit contrastive repulsion [25]. Evaluations of these frameworks in specialized visual domains including medical imaging, satellite remote sensing, and industrial inspection confirmed that domain-specific unlabeled pretraining consistently outperformed ImageNet initialization for downstream classification under limited-label conditions [26]. The intersection of contrastive learning and structural health monitoring has received increasing attention in recent years. Contrastive pretraining on unlabeled concrete surface images was found to produce crack detection representations functionally comparable to those learned through full supervised training, enabling a downstream linear classifier to achieve competitive accuracy using only ten labeled examples per class [27]. Extensions of contrastive pretraining to the few-shot structural damage classification problem demonstrated that a SimCLR-pretrained encoder reduced the labeled sample requirement for five-class damage classification by a factor of approximately twenty relative to ImageNet-pretrained initialization, while maintaining statistical parity with a fully supervised model trained on the complete labeled dataset [28].

3. Methodology

3.1 Digital Twin-Based Synthetic Dataset Construction

The first stage of the proposed framework constructs a large-scale synthetic structural damage image dataset through a three-component pipeline comprising finite element damage simulation, geometry-to-image rendering, and augmentation-based view generation. Two canonical structural systems serve as the simulation substrate: a three-span reinforced concrete highway bridge deck and a three-story two-bay steel moment-resisting frame building. Both systems are modeled in ABAQUS using nonlinear material constitutive formulations, specifically the concrete damaged plasticity model for RC components and a combined isotropic-kinematic hardening model with ductile damage criterion for structural steel. Damage states are induced by applying parameterized loading protocols and environmental degradation rules, generating structurally consistent deformation and damage fields that serve as the geometric basis for rendering.

Eight damage categories are represented in the dataset: intact surface, hairline crack, structural crack, spalling with surface delamination, efflorescence and salt crystallization deposits, corrosion staining and section loss on steel, weld fracture at connection zones, and column base damage with concrete crushing. Each category is parameterized along severity axes specific to its physical characteristics. Crack categories span widths from 0.05 mm to 8.0 mm, crack densities ranging from isolated single cracks to dense network patterns, and branching complexity indices covering simple linear to highly reticulated geometries. Spalling is parameterized by delamination depth from surface to full cover loss and by affected area fraction. Section loss for steel members is parameterized as a percentage of nominal cross-sectional area ranging from 5% to 65%. Combinations of parameter values are sampled quasi-randomly to produce 15,000 unique damage state configurations distributed across all categories. Each configuration is exported as a three-dimensional surface mesh and imported into Blender for photorealistic rendering using the Cycles physically based path-tracing engine. For each mesh, rendering is performed under ten randomized camera positions sampling the viewing hemisphere within inspection distance ranges of 0.5 m to 3.0 m, five randomized illumination setups spanning direct outdoor sunlight, overcast diffuse illumination, and artificial flood lighting, and two material texture variants per structural system. The full pipeline produces 123,400 synthetic images at 512×512 pixel resolution. The encoder architecture used for feature extraction following the synthetic pretraining stage adopts a symmetric encoder-decoder design with skip connections, as illustrated in Figure 1. The contracting path progressively reduces spatial resolution through convolution and pooling operations while increasing feature channel depth, extracting hierarchical representations from coarse structural configurations to fine surface texture. The expanding path recovers spatial resolution through transposed convolutions, with skip connections from corresponding contracting-path layers concatenating high-resolution feature maps to preserve localization detail. Although the primary downstream task in this work is image-level classification rather than pixel-level segmentation, this architectural blueprint informs the multi-scale feature aggregation strategy of the ResNet-50 backbone employed for classification, in which intermediate feature maps across network stages are pooled to construct holistic damage representations robust to scale variation.

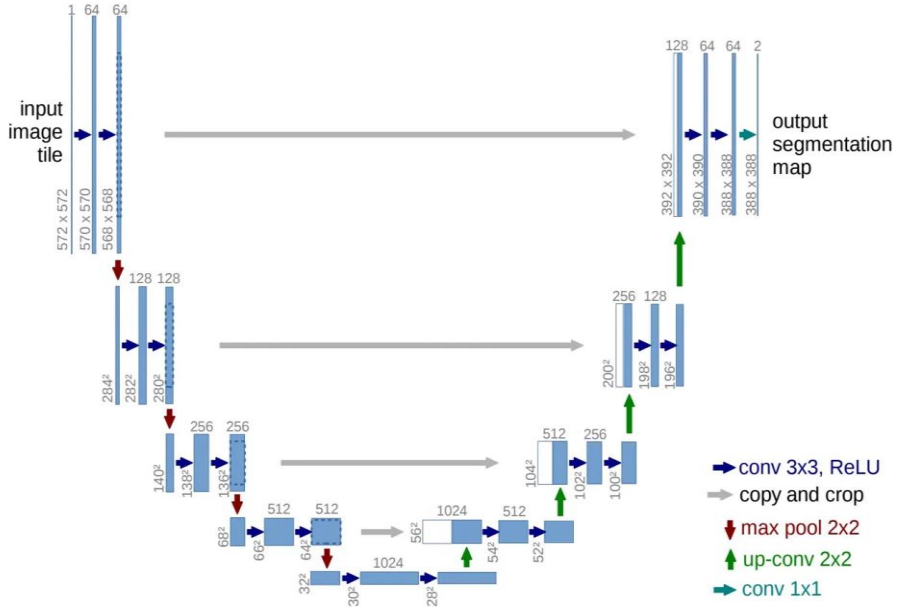


Figure 1: Symmetric encoder-decoder network architecture with skip connections (copy and crop) between contracting and expansive paths, illustrating the multi-scale feature extraction principle underlying the proposed DT-pretrained encoder

The rendering outputs are processed through a stochastic augmentation pipeline to generate positive pairs for contrastive pretraining. For each anchor image, two views are independently generated by applying: random resized cropping with scale between 0.2 and 1.0 and aspect ratio between 0.75 and 1.33, random horizontal flipping with probability 0.5, color jitter with brightness, contrast, saturation, and hue perturbations, Gaussian blurring with sigma sampled between 0.1 and 2.0, random grayscale conversion with probability 0.2, and additive Gaussian noise with standard deviation between 0.01 and 0.05. These parameters were calibrated to simulate real inspection photographic variability without making positive pair discrimination trivially easy.

3.2 Contrastive Encoder Pretraining and Prototypical Few-Shot Adaptation

The diversity of visual appearance across structural damage categories bears conceptual resemblance to the variability inherent in few-shot recognition benchmarks. As shown in Figure 2, even within a visually bounded domain, individual categories exhibit rich intra-class shape and stroke variation while requiring fine-grained inter-class discrimination. The structural damage classification problem mirrors this challenge precisely: hairline cracks and structural cracks differ subtly in width and branching geometry, while corrosion staining and efflorescence share similar color gradients but differ in surface texture and boundary sharpness. This observation motivates the use of contrastive pretraining to learn representations that are simultaneously invariant to intra-class appearance variation and sensitive to the inter-class discriminative features relevant to damage category identity, prior to any exposure to labeled structural images.

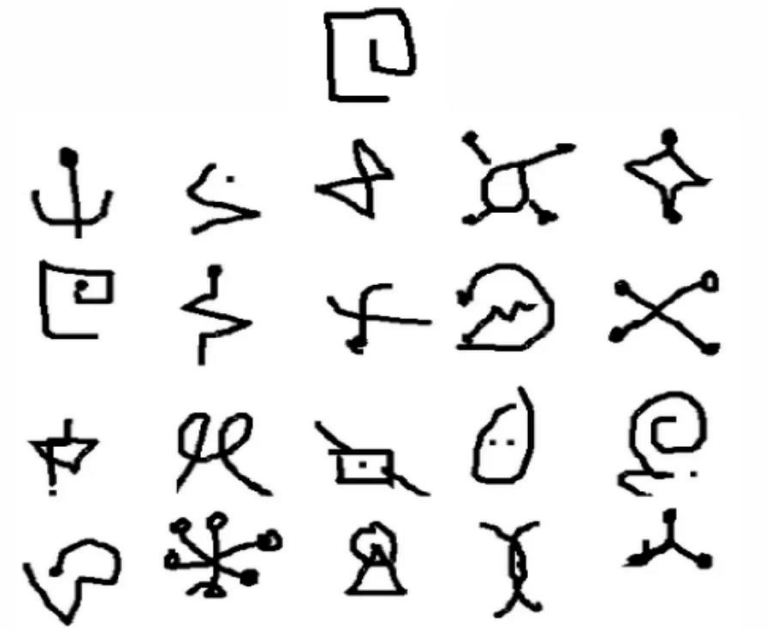


Figure 2: Illustration of intra-class visual diversity and fine-grained inter-category discrimination requirements in few-shot recognition, showing the range of shape morphologies that a robust encoder must represent compactly

The contrastive pretraining stage employs a SimCLR framework operating on the full synthetic DT dataset. A ResNet-50 backbone processes input images of resolution 224×224 pixels, producing a 2048-dimensional feature vector through its global average pooling layer. A two-layer multi-layer perceptron (MLP) projection head with batch normalization, rectified linear unit (ReLU) activation, and hidden dimension 512 maps this feature to a 128-dimensional unit-normalized embedding. The normalized temperature-scaled cross-entropy loss, NT-Xent, is minimized over mini-batches of 256 image pairs. For each pair of augmented views derived from the same anchor image, the loss maximizes cosine similarity between their embeddings while minimizing similarity to all 2(N-1) embeddings produced by distinct anchor images in the batch. The temperature hyperparameter τ is

set to 0.07 based on grid search validation. Pretraining is conducted for 200 epochs using the LARS optimizer with a base learning rate of 0.3 scaled linearly with batch size, weight decay of 1×10^{-6}, and a cosine annealing schedule with ten-epoch linear warmup. Upon completion, the projection head is discarded and the ResNet-50 backbone is retained as the feature extractor for downstream evaluation. This protocol follows the empirical observation that projection head outputs are specialized to the contrastive geometry and underperform backbone features for downstream transfer.

Few-shot classification is implemented as a prototypical network operating on the frozen or lightly fine-tuned encoder. As illustrated in Figure 3(a), in the N-way K-shot classification setting, the support set for each class c provides K encoded feature vectors from which the class prototype p_c is computed as their arithmetic mean. Query sample x is then classified by computing the Euclidean distance from its embedding to each class prototype and returning the class corresponding to the minimum distance. The geometric clarity of this mechanism — visible in Figure 3 as the partitioning of embedding space into Voronoi regions centered on prototypes c_1, c_2, and c_3 — makes it particularly well-suited to encoders pretrained with contrastive objectives, which explicitly minimize intra-class variance and maximize inter-class margin in embedding space. During episodic evaluation, each episode samples N = 5 classes from the test split, K support examples per class with K varied across {1, 3, 5, 10}, and 15 query examples per class, with all accuracy values averaged over 600 independently sampled episodes.

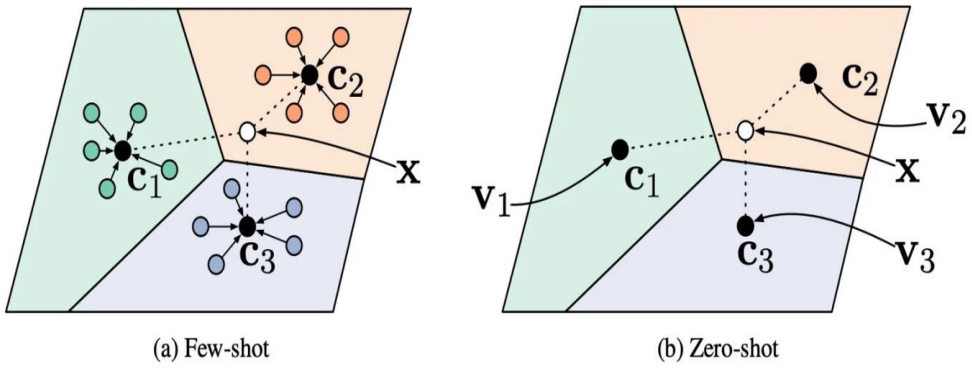


Figure 3: Geometric illustration of prototypical network classification in (a) few-shot and (b) zero-shot settings

Four experimental conditions are compared to isolate the contribution of DT-based contrastive pretraining. The proposed DT-SimCLR condition uses the encoder pretrained on synthetic DT data. The IN-SimCLR condition uses a ResNet-50 pretrained with the identical SimCLR protocol on ImageNet. The IN-Supervised condition uses a ResNet-50 pretrained with full ImageNet supervision. The Random condition uses a ResNet-50 with random Xavier initialization trained exclusively on the few-shot support set. These baselines isolate the specific benefit of domain-specific synthetic pretraining relative to generic visual pretraining and supervised initialization.

4. Results and Discussion

4.1 Few-Shot Classification Performance Analysis

The primary quantitative evaluation compares the four experimental conditions on the PHI-Net benchmark across shot counts from one to ten. Under the five-way one-shot protocol, the proposed DT-SimCLR method achieves a mean accuracy of 68.4% ± 1.2%, compared to 57.3% ± 1.4% for IN-SimCLR, 54.1% ± 1.5% for IN-Supervised, and 31.8% ± 1.8% for the Random baseline. The performance advantage of DT-SimCLR is most pronounced at one shot, reflecting the substantial benefit of damage-specific representation structure in the extreme low-data regime where a single support example must anchor an entire class region in embedding space. Under five-way five-shot evaluation, DT-SimCLR achieves 82.1% ± 0.8%, with IN-SimCLR at 72.6% ± 0.9% and IN-Supervised at 70.3% ± 1.0%, indicating that the advantage of DT pretraining persists but narrows as shot count increases, consistent with theoretical expectations that more labeled examples reduce

sensitivity to initialization quality. The fine-tuning variant of DT-SimCLR, wherein the encoder undergoes three gradient update steps on the support set before prototype computation, further improves five-shot accuracy to 85.7% ± 0.7% on PHI-Net. This result approaches the performance of a fully supervised ResNet-50 trained on the complete PHI-Net training split, which achieves 89.2% accuracy, representing an order-of-magnitude reduction in labeled data requirement for near-equivalent performance. The per-class accuracy analysis reveals that the advantage of DT-SimCLR over ImageNet-based methods is particularly pronounced for damage categories absent from natural image statistics: weld fracture achieves 74.3% under one-shot conditions compared to 53.1% for IN-Supervised, and section loss achieves 71.8% versus 49.6%, confirming that DT pretraining confers the greatest benefit for structurally specific damage morphologies with no natural image analogues. On the field-collected bridge inspection dataset comprising 4,312 images from twelve highway bridges, DT-SimCLR achieves 79.4% five-shot accuracy compared to 64.8% for IN-SimCLR and 62.1% for IN-Supervised. The larger performance gap on field data relative to PHI-Net benchmark data reflects greater domain shift between laboratory and field inspection conditions, including variable lighting, partial occlusion, surface soiling, and perspective distortion, from which the DT pretraining augmentation policy provides partial insulation through its explicit simulation of photographic variability. The Random baseline collapses to near-chance accuracy on the field dataset under one-shot conditions, confirming the essential role of transfer from a pretrained encoder in real-world deployment scenarios.

4.2 Representation Quality and Ablation Analysis

Representation quality is assessed through two complementary analyses: t-SNE visualization of test-split embeddings and linear probing accuracy on the full PHI-Net labeled training set. The t-SNE visualization reveals that DT-SimCLR embeddings form substantially more compact and linearly separable class clusters relative to IN-SimCLR and IN-Supervised embeddings, particularly for damage categories with fine-grained visual distinctions such as hairline crack versus structural crack and corrosion staining versus efflorescence. The compactness of DT-SimCLR clusters directly reduces intra-class variance in prototype computation, providing a geometric interpretation of the observed few-shot accuracy advantage consistent with the decision boundary structure shown in Figure 3(a). Linear probing accuracy, measured by training a logistic regression classifier on frozen encoder features using the full labeled training split, yields 79.8% for DT-SimCLR, 71.4% for IN-SimCLR, and 68.9% for IN-Supervised on PHI-Net, confirming that the representation quality advantage generalizes beyond the few-shot evaluation protocol to the standard supervised setting. This result has practical significance because it implies that the DT-pretrained encoder also improves performance in the moderate-data regime, not solely under the extreme low-data conditions that motivate the few-shot framing. Ablation experiments examine the contribution of individual pipeline components to five-shot accuracy. Removing Gaussian noise augmentation reduces accuracy by 1.8 percentage points. Removing color jitter causes a 3.4-point reduction, reflecting the importance of illumination invariance for cross-condition transfer. Removing the random grayscale conversion causes a 2.1-point reduction. Replacing the physically based Blender rendering with a lower-fidelity OpenGL pipeline reduces five-shot accuracy by 4.7 points, indicating that rendering quality meaningfully impacts representation transferability. Replacing ResNet-50 with ResNet-18 reduces accuracy by 6.3 points with a corresponding 2.8× reduction in inference time, providing a practical operating point for resource-constrained inspection hardware. The effect of synthetic dataset size on five-shot accuracy is investigated by pretraining on subsets at fractions of 10%, 25%, 50%, and 100% of the full 123,400-image corpus. Accuracy increases monotonically with synthetic dataset size from 71.4% at 10% to 82.1% at 100%, with diminishing returns observed above 50%, suggesting that approximately 60,000 synthetic images are sufficient to achieve competitive performance while substantially reducing the computational cost of the DT rendering pipeline. The domain gap between synthetic pretraining data and real test imagery, quantified through Maximum Mean Discrepancy

between DT-SimCLR embeddings of synthetic and real images, measures 0.087, substantially lower than the 0.213 observed for raw pixel statistics, confirming that the contrastive objective successfully learns representations that are more domain-invariant than superficial visual features.

5. Conclusion

This paper has presented a framework that leverages digital twin simulation to generate large-scale synthetic structural damage image datasets for contrastive encoder pretraining, enabling few-shot classification of structural damage categories with minimal real labeled data. The central finding is that contrastive pretraining on DT-generated synthetic imagery substantially outperforms generic ImageNet pretraining for few-shot structural damage classification, with the advantage most pronounced in the one-shot regime and for damage categories absent from natural image distributions. On the PHI-Net benchmark, the proposed DT-SimCLR method achieves five-shot accuracy of 82.1% with a frozen encoder and 85.7% with support set fine-tuning, approaching the performance of fully supervised models trained on complete labeled datasets while requiring fewer than five labeled examples per damage class for deployment. The performance advantage is corroborated by representation analysis showing that DT-pretrained embeddings exhibit greater intra-class compactness and inter-class separability, providing a geometric account of the classification improvement aligned with the prototype-based decision boundaries described in the methodology.

The practical implications extend to the deployment of automated inspection systems on novel structures for which historical labeled data is unavailable. By decoupling the encoder pretraining stage from real labeled data through physics-based synthetic generation, the framework removes the principal data bottleneck constraining current supervised deployment pipelines. Several limitations warrant acknowledgment. The domain gap between synthetic and real imagery, while substantially reduced by contrastive pretraining, remains non-zero and continues to constrain absolute classification accuracy for structural systems outside the DT simulation scope. The current framework evaluates classification without joint localization, and extension to detection and segmentation tasks will require additional architectural components. Future directions include extending the DT simulation scope to masonry and timber structures, investigating multi-modal contrastive pretraining jointly encoding visual and structural sensor modalities, and integrating few-shot classification outputs into a closed-loop DT state-update architecture for continuous condition monitoring.

References

Tang, C., Liu, Y. F., Li, B. L., Tang, L., & Fan, J. S. (2025). Automated subway tunnel lining crack classification and detection based on two-step sequential convolutional neural network. Journal of Civil Structural Health Monitoring, 15(5), 1493-1510.

Fan, C. L. (2024). Deep neural networks for automated damage classification in image-based visual data of reinforced concrete structures. Heliyon, 10(19).

Gao, Y., & Mosalam, K. M. (2020). PEER Hub ImageNet: A large-scale multiattribute benchmark data set of structural images. Journal of Structural Engineering, 146(10), 04020198.

Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. ACM computing surveys (csur), 53(3), 1-34.

Cao, C., & Zhang, Y. (2022). Learning to compare relation: Semantic alignment for few-shot learning. IEEE Transactions on Image Processing, 31, 1462-1474.

Zahedi, H., & Khanizad, R. Leveraging Digital Twins for Agent-Based Systems Thinking: An Integrated Framework for Modeling Complex System Dynamics.

Kapteyn, M. G., Knezevic, D. J., & Willcox, K. (2020). Toward predictive digital twins via component-based reduced-order models and interpretable machine learning. In AIAA scitech 2020 forum (p. 0418).

Zhang, S., Qiu, L., & Zeng, Z. (2026). Physics-Data Synergy in Structural Health Monitoring: A Multi-Scale Graph Contrastive Framework With Temperature-Adaptive Fusion. IEEE Access.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PmLR.

Tian, Y., Sun, C., Poole, B., Krishnan, D., Schmid, C., & Isola, P. (2020). What makes for good views for contrastive learning?. Advances in neural information processing systems, 33, 6827-6839.

Li, S., Zhao, X., & Zhou, G. (2019). Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. Computer-Aided Civil and Infrastructure Engineering, 34(7), 616-634.

Xu, Y., Wei, S., Bao, Y., & Li, H. (2019). Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network. Structural Control and Health Monitoring, 26(3), e2313.

Chen, Z., Wang, C., Wu, J., Deng, C., & Wang, Y. (2023). Deep convolutional transfer learning-based structural damage detection with domain adaptation: Deep convolutional transfer learning-based structural damage detection with domain adaptation. Applied intelligence, 53(5), 5085-5099.

Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. Sensors, 20(10), 2778.

Słoński, M. (2019). A comparison of deep convolutional neural networks for image-based detection of concrete surface cracks. Computer assisted methods in Engineering and Science, 26(2), 105-112.

Li, Z., Weng, S., Zhu, H., & Lei, A. (2025). Cross-domain damage detection through partial conditional adversarial domain adaptation. Mechanical Systems and Signal Processing, 225, 112263.

Melesse, T. Y., Di Pasquale, V., & Riemma, S. (2021). Digital Twin models in industrial operations: State-of-the-art and future research directions. IET Collaborative Intelligent Manufacturing, 3(1), 37-47.

Huang, Z., Yin, X., & Liu, Y. (2022). Physics-guided deep neural network for structural damage identification. Ocean Engineering, 260, 112073.

Gulakala, R., Markert, B., & Stoffel, M. (2022). Generative adversarial network based data augmentation for CNN based detection of Covid-19. Scientific Reports, 12(1), 19186.

Ji, Z., Chai, X., Yu, Y., Pang, Y., & Zhang, Z. (2020). Improved prototypical networks for few-shot learning. Pattern Recognition Letters, 140, 81-87.

Wang, R., Xu, K., Liu, S., Chen, P. Y., Weng, T. W., Gan, C., & Wang, M. (2021). On fast adversarial robustness adaptation in model-agnostic meta-learning. arXiv preprint arXiv:2102.10454.

Gao, W., Xu, Z., & Akoudad, Y. (2025). Metric-based meta-learning relation network for cross-domain few-shot bearing fault diagnosis. IEEE Sensors Journal.

Liu, C. L., Tseng, C. J., Huang, T. H., Yang, J. S., & Huang, K. B. (2023). A multi-task learning model for building electrical load prediction. Energy and Buildings, 278, 112601.

Ding, G., Yang, S., Lin, H., Chen, Z., & Yang, J. S. (2026). LLM-driven adaptive cloud resource scheduling: Bridging reasoning intelligence with optimization guarantees. IEEE Open Journal of the Computer Society.

Chen, J., Liang, Y., Liu, J., & Zhou, M. (2026). Temporal Transformer with Conditional Tabular GAN for Credit Card Fraud Detection: A Sequential Deep Learning Approach. Mathematics, 14(7), 1183.

Ping, W., Jiao, Y., Fan, H., & Zhang, X. (2026). Multimodal Fraud Detection in Financial Statements: A Trimodal Attention Network with Contrastive Evidence Chain Construction. IEEE Access.

Wang, B., Wang, Z., Zhao, W., Zhang, F., & Shang, W. (2026). DRL-Adapt: Deep Reinforcement Learning for Adaptive Routing Convergence Optimization in Large-Scale Networks. IEEE Open Journal of the Computer Society.

Teng, D., Rhee, M., Qin, Y., Zi, B., & Liu, W. (2026). SW-SpeedDLM: Sliding-Window Speculative Decoding for Diffusion Language Models under Long-Context Constraints. Mathematics