International Journal of Geo-Information

MDPI

Article

Semi-Supervised AI for Architectural Heritage Classification and Style Lineage Discovery in Chinese Traditional Settlements

Qing Han [1,2], Zicheng Wang [3], Chao Yin [4,*], Zhiwei Hou [5] and Tianci Yao [4]

1 National-Local Joint Engineering Laboratory on Digital Preservation and Innovative Technologies for the Culture of Traditional Villages and Towns, Hengyang Normal University, Hengyang 421002, China; ihqing@hynu.edu.cn
2 College of Geography and Tourism, Hengyang Normal University, Hengyang 421002, China
3 The Academy of Digital China, Fuzhou University, Fuzhou 350108, China; wangzicheng@fzu.edu.cn
4 Guangzhou Institute of Geography, Guangdong Academy of Sciences, Guangzhou 510070, China; yaotc.17b@igsnrr.ac.cn
5 Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 510070, China; houzw@lreis.ac.cn
* Correspondence: cyinac@connect.ust.hk

Abstract

Large-scale classification of architectural styles in Chinese traditional settlements is important for heritage conservation and geospatial documentation, but scalable deployment remains constrained by the high cost of expert annotation because villages are widely distributed, the imagery is captured from heterogeneous viewpoints, and each architectural tradition exhibits substantial intra-class variation. To address this bottleneck, we propose CTSMatch, a label-efficient semi-supervised framework that combines an ImageNet-pretrained EfficientNetV2 backbone with SoftMatch-based adaptive pseudo-label weighting so that ambiguous but informative unlabeled samples can still contribute to training, thereby reducing reliance on costly expert annotation. We also construct SemiCTS, an extension of the original CTS dataset that adds 4360 unlabeled images. Using only 545 labeled samples, CTSMatch achieves 96.93% accuracy on SemiCTS, outperforming the strongest fully supervised baseline (Dense-TL-Aug) by 2.73 percentage points and two standard semi-supervised baselines (FixMatch and FreeMatch) by 3.06 percentage points. Beyond classification, we further analyze the feature space to examine stylistic lineage through intra-style heterogeneity, inter-style transitions, and outlier detection. The results reveal two broad regional groupings, a northern cluster (Jing, Jin, Su) and a southern cluster (Chuan, Min, Wan), connected by gradual transitions rather than rigid boundaries. Approximately 15% of the samples are identified as atypical cases, including 8.7% comprising regional variants and 6.3% comprising hybrid forms. These findings show that CTSMatch provides a practical label-efficient framework for architectural heritage classification while supporting the interpretable analysis of stylistic diversification and convergence in Chinese traditional settlements.

Keywords: architectural heritage classification; Chinese traditional settlements; semi-supervised learning; cultural heritage visualization; architectural style lineage discovery

Check for updates

Academic Editor: Wolfgang Kainz, Ivana Ivánová, Yongze Song and Sisi Zlatanova

Received: 29 March 2026
Revised: 7 May 2026
Accepted: 15 May 2026
Published: 20 May 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Traditional villages preserve China's agrarian civilization, vernacular architecture, indigenous knowledge, and community-based social structures [1]. Rapid urbanization is placing these heritage sites under increasing pressure, making systematic documentation

and conservation urgent. Architectural styles encode region-specific historical, cultural, and environmental knowledge [2,3]. Reliable style classification therefore supports heritage research, conservation planning, rural revitalization, and tourism development [4–6]. In China, where diverse traditional architectures extend across thousands of settlements, automated style classification can also support place-sensitive governance and digital heritage management [7].

Architectural style classification has traditionally relied on manual analysis, which demands substantial expert knowledge [7,8]. Traditional typological and facade-based studies remain essential for documenting heritage forms, but they still depend mainly on expert qualitative judgment rather than scalable image learning [9]. Over the past decade, deep learning and transfer learning have substantially improved architectural image classification [10–14]. In the related area of cultural heritage image classification, studies of pretrained CNNs and comparative transfer learning pipelines likewise support the use of transfer learning for visually heterogeneous heritage datasets [15,16]. Xu et al. [12] introduced an early learning-based approach, and later studies used convolutional neural networks to extract style-relevant visual features [13]. For Chinese traditional settlements (CTS), Han et al. [14] established the first dedicated dataset and proposed a fully supervised framework that combines transfer learning with automatic augmentation. At a related regional scale, Zou et al. [17] used machine learning to quantify facade-form features in Hubei architectural heritage, showing that image-based learning can recover geographically differentiated architectural traits. From a geoinformation perspective, the territorial distribution of architectural styles remains an important complementary dimension of heritage interpretation. Prior CTS-related work has examined the spatial distribution of traditional Chinese architectural styles at multiple scales [18], while recent geomatics research shows that satellite remote sensing and deep learning can complement, rather than replace, ground-image-based heritage analysis in broader preservation planning [19]. These studies demonstrate that deep learning can recognize architectural styles effectively when labeled data are available, but they do not remove the practical bottleneck of building large, reliable annotated datasets for geographically dispersed traditional settlements.

However, applying such methods at scale to CTS remains difficult for three main reasons. First, traditional settlements are widely distributed across remote regions of China, so image collection requires substantial travel, access coordination, and time. Second, CTS imagery exhibits strong multi-view heterogeneity: the same style may appear quite differently in frontal views, courtyard perspectives, interior details, and settlement-scale scenes. Third, each architectural tradition contains substantial intra-class variation across sub-regions, historical periods, and building functions. Together, these factors make expert labeling expensive and limit the scalability of purely supervised learning. Semi-supervised learning offers an alternative because unlabeled architectural imagery can be collected more scalably from public sources while reducing dependence on expert labels. However, a key gap remains: existing semi-supervised methods have not been systematically evaluated on architectural style data, and their effectiveness under the weak-style conditions of CTS imagery is still unclear.

To address this gap, we propose CTSMatch, a label-efficient semi-supervised framework that combines transfer learning with SoftMatch's adaptive pseudo-label weighting [20]. In addition to improving classification under limited annotation, we further analyze the feature space to examine stylistic continuity, transition zones, and atypical cases in Chinese traditional settlements. We make four contributions:

- **SemiCTS benchmark:** We extend the original CTS dataset with 4360 unlabeled images, creating a benchmark for label-efficient architectural heritage classification under realistic annotation scarcity.

- **Weak-style-aware CTSMatch framework:** We combine ImageNet transfer learning with SoftMatch's adaptive pseudo-label weighting so that ambiguous but informative unlabeled samples can contribute to training instead of being discarded by hard thresholds, yielding 96.93% accuracy with only 545 labeled images.
- **Quantitative heterogeneity analysis:** We measure intra-style variation across six architectural traditions, providing quantitative support for weak-style analysis across regional typologies and viewing conditions.
- **Stylistic lineage discovery:** Building on this feature-space evidence, we identify regional clusters, gradual transitions, regional variants, and hybrid forms, revealing a northern cluster (Jing, Jin, Su) and a southern cluster (Chuan, Min, Wan).

The remainder of this paper is organized as follows. Section 2 reviews architectural classification and semi-supervised learning studies. Section 3 describes the SemiCTS dataset and CTSMatch framework. Section 4 presents experimental results comparing CTSMatch with baseline methods. Section 5 analyzes architectural style lineage through feature-space visualization and heterogeneity metrics. Section 6 discusses the findings in the context of recent studies, and Section 7 concludes this paper.

2. Literature Review

2.1. Deep Learning for Chinese Traditional Architecture Classification

Traditional architecture reflects regional cultural and historical identity and is shaped by religious, social, and economic conditions [3]. Early studies of traditional villages relied on Geographic Information Systems (GIS) and spatial analysis to document historical and cultural patterns [21–23]. More recently, computer-vision methods have used convolutional neural networks (CNNs) to learn architectural features directly from imagery [24–26].

Zhao et al. [24] were among the first to apply deep CNNs to architectural style classification. Gao et al. [25] and Wang et al. [26] further improved classification with attention mechanisms and refined network architectures. For CTS, Han et al. [14] established the first dedicated dataset and proposed Dense-TL-Aug, a fully supervised framework that combines transfer learning with automatic augmentation. Importantly, that work also used Grad-CAM to visualize the architectural components (e.g., roofs, walls, columns, and ornamental details) that the CNN relies on for each style, showing that the model attends to diagnostically meaningful features rather than spurious environmental patterns. The same research line later analyzed the spatial distribution of traditional Chinese architectural styles at multiple scales [18].

These studies show that deep learning recognizes CTS styles effectively when labeled data are abundant. However, fully supervised approaches require extensive manual annotation, which limits practical deployment in heritage documentation where expert labeling is expensive. While architectural refinements, pretraining, and stronger augmentation improve model capacity, they do not address the label-scarcity constraint. Semi-supervised learning provides a solution by using abundant unlabeled imagery to supplement limited labeled examples, allowing accurate CTS classification without comprehensive annotation.

2.2. Semi-Supervised Deep Learning Methods

Semi-supervised deep learning uses both labeled and unlabeled data during training, balancing supervised and unsupervised learning [27,28]. This approach is valuable for architectural style classification of Chinese traditional settlements, where labeled data are scarce and expensive. Semi-supervised methods work in domains with similar labeling constraints, including medicine, engineering, natural language processing, and remote sensing [29–32], with some recent advancements surpassing fully supervised baselines.

Semi-supervised deep learning methods include generative, consistency-regularization, graph-based, pseudo-labeling, and hybrid approaches [27]. Hybrid approaches integrate complementary techniques. MixMatch [33] pioneered this by combining consistency regularization, entropy minimization, and traditional regularization. FixMatch [34] streamlined the approach with a fixed confidence threshold for pseudo-label filtering, but this rigid threshold often discards potentially valuable low-confidence predictions. FlexMatch [35] and FreeMatch [36] introduced adaptive thresholds that evolve during training, allowing models to learn from more pseudo-labels during early stages. SoftMatch [20] addressed the quantity–quality trade-off by weighting loss according to pseudo-label confidence distribution, which works well for CTS classification where confident predictions emerge gradually as the model learns regional and stylistic variations.

2.3. Weak-Style Classification Challenges

Style classification is challenging because of complex and diverse semantic features in real-world style images. While deep learning has advanced this field, difficulties remain with subtle inter-class differences and incomplete semantic representation in datasets [37,38].

Jiang et al. [39] described this as a "weak-style classification problem", in which blurred class boundaries arise from incomplete and heterogeneous semantic cues. They proposed the Consensus Style Centralizing Auto-Encoder (CSCAE) method, which learns multiple auto-encoders under consistency constraints. A subsequent refinement introduced a linear CSCAE for computational efficiency and used dimensionality reduction for visualization [40]. With the rise of CNNs, however, comparable solutions for weak-style classification within modern deep learning pipelines remain limited.

The weak-style problem is especially pronounced in CTS, where architectural traditions vary across viewpoints, settlement scales, sub-regions, and building functions. Because traditional Chinese architecture evolved through gradual geographical and historical transitions rather than explicit taxonomic rules, many samples contain blended cues that even experts struggle to classify consistently. This ambiguity limits supervised learning because sparse labels cannot represent the full range of stylistic expression. Our semi-supervised framework addresses this challenge by leveraging limited labels and abundant unlabeled imagery to capture stylistic continuity and lineage relationships.

3. Methodology

This section describes our methodology for label-efficient CTS classification and stylistic lineage discovery. As shown in Figure 1, we implement a three-stage pipeline: (1) SemiCTS dataset construction, extending the original CTS dataset [14] with unlabeled samples; (2) CTSMatch framework; and (3) feature-space analysis to find architectural style relationships through clustering and outlier detection.

Figure 1 should therefore be read as a progression from data construction to weak-style-aware training and finally to post hoc heritage interpretation: unlabeled data improve classification performance and also shape an embedding space that can later be analyzed for stylistic continuity, cluster structure, and atypical cases.

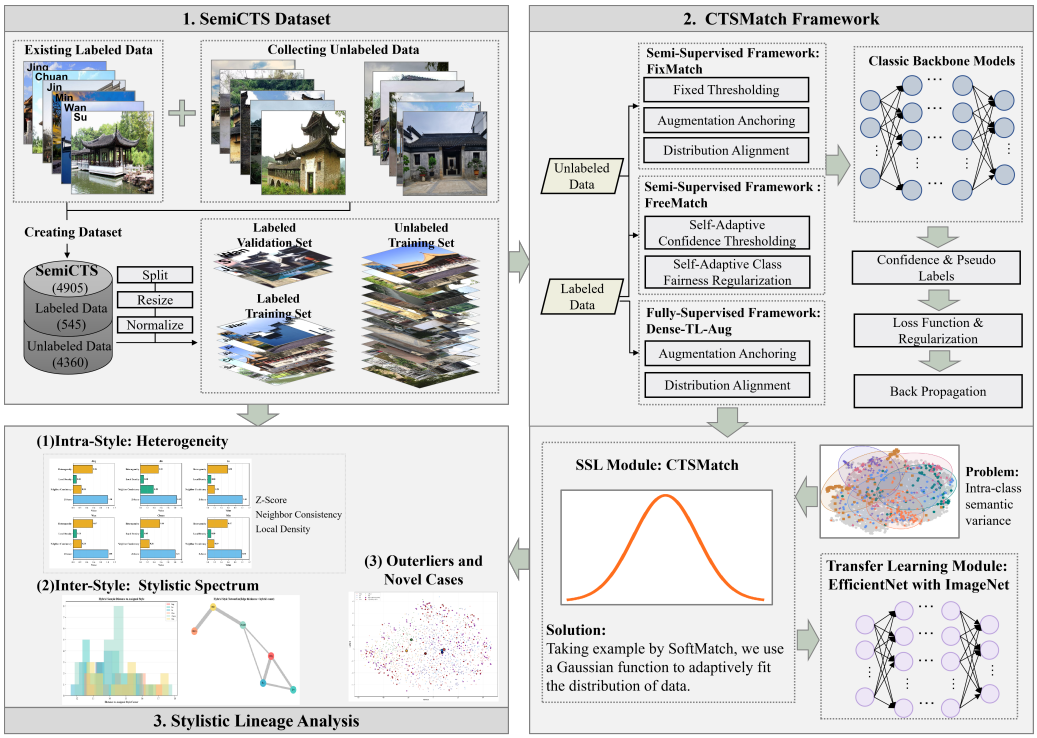

Figure 1. Semi-supervised pipeline for Chinese traditional settlement (CTS) style classification. The workflow consists of three stages: (1) SemiCTS dataset construction with labeled and unlabeled images, (2) CTSMatch framework training with adaptive pseudo-label weighting, and (3) performance evaluation and stylistic lineage analysis through feature-space visualization.

3.1. SemiCTS Dataset Construction

China contains hundreds of thousands of traditional settlements. We focus on six representative architectural traditions: Chuan, Jin, Jing, Min, Su, and Wan. Our study area covers the first three batches of Chinese traditional villages published by the Ministry of Housing and Urban–Rural Development, encompassing more than 2500 settlements. Figure 2 visualizes their geographic distribution using location data from the *Digital Journal of Global Change Data* Repository [41]. The labeled architectural image data derive from the CTS dataset released by Han et al. [14] and hosted at https://github.com/PointCloudYC/CTS (accessed on 20 May 2025).

Despite its value for fully supervised learning, the CTS dataset exhibits the weak-style problem described in Section 1. First, semantic coverage is incomplete and heterogeneous. Images collected from varied and often partial viewpoints—frontal views, interior details, and compound-scale scenes—do not consistently expose the defining stylistic cues (Table 1). Second, intra-class variation is substantial. Within a single style such as Jin, visual forms range from vernacular cave dwellings on the Loess Plateau to ornate merchant courtyards. Third, limited labeled data constrains semi-supervised learning. The original 1226 labeled images provide sparse supervision for models that must learn fine-grained stylistic cues.

Additional unlabeled images were systematically collected from publicly accessible web sources to expand the dataset. To improve representativeness and reduce selection bias, we used a village-level stratified sampling strategy based on the first three batches of the Chinese Traditional Villages Catalogue published by the Ministry of Housing and Urban–Rural Development. For each catalogued village, we collected multiple architectural images spanning building types, viewpoints, and settlement scales. This design allowed the unlabeled subset to capture both within-village variation and between-village diversity rather than over-representing a small number of well-documented settlements. No style labels were assigned during collection; all images entered the unlabeled pool without

architectural-style annotations, and pseudo-labels were generated only during training. This procedure broadens semantic coverage while preserving data provenance.

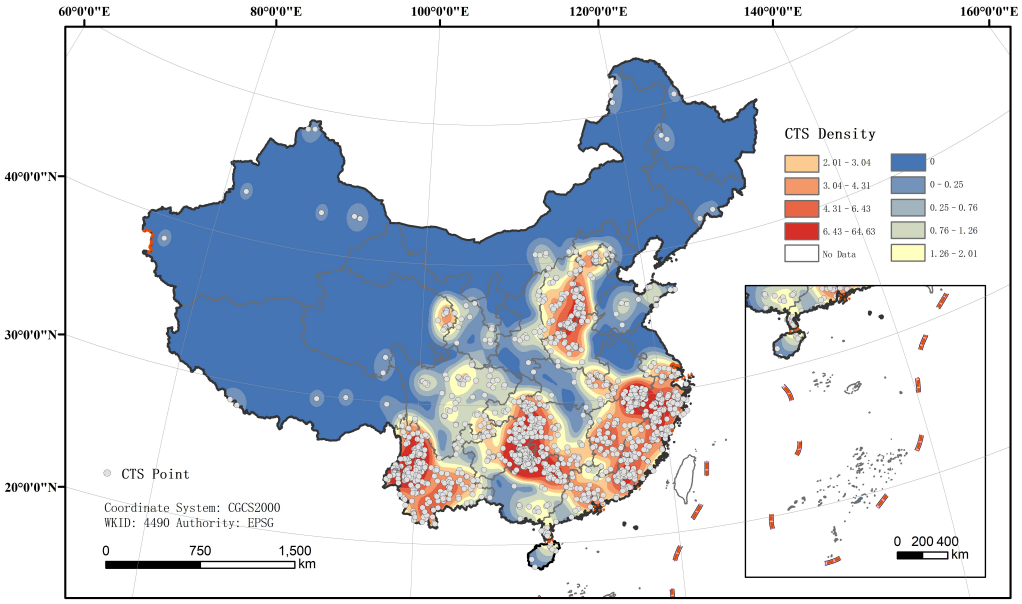

Figure 2. Geographical distribution of Chinese traditional settlements. The density map shows the first three batches of Chinese traditional villages, highlighting the regions where the six architectural styles are predominantly distributed.

Table 1. Multi-view heterogeneity in CTS imagery. Using Jin style as an example, the table illustrates how the same architectural tradition appears across different architectural types (courtyard architecture, cave dwelling), viewing perspectives (front view, interior details), and settlement scales (individual building vs. group view). This visual variability exemplifies the weak-style classification problem.

Style	Architecture Type	Partial View		Group View
		Front View	Inside Details	
Jin	Courtyard architecture			
	Cave dwelling architecture			

The collected images were then manually reviewed for quality control. We removed non-architectural images (e.g., pure landscapes and text pages), excluded modern structures outside the traditional architectural context, eliminated duplicate or near-duplicate images, and discarded images with severe quality problems such as extreme blurriness or heavy watermarking. After cleaning, 4360 unlabeled images were retained for training. All images were resized to 224 × 224 pixels and normalized before training. Original image resolutions vary across web sources and are therefore not uniform. Because the evaluated

models already incorporate their own augmentation strategies, we applied no additional preprocessing augmentation.

For the labeled subset, architectural heritage experts reviewed the original CTS dataset's 1226 annotated images and selected 545 images with clear, unambiguous stylistic features representative of each tradition. Images with ambiguous style cues, mixed architectural elements, or poor quality were excluded. This expert-curated subset simulates a realistic low-label scenario in which only high-confidence annotations are available. The 545 labeled images were split into training and validation sets at a 70:30 ratio. Table 2 summarizes the SemiCTS dataset.

Table 2. Comparison of data size between the original CTS dataset and the proposed SemiCTS dataset. Each image comes from a distinct traditional village in the first three batches, with more than three images per village, ensuring spatial and typological diversity.

Dataset	Chuan	Jin	Jing	Min	Su	Wan	Unlab.	Total
Original CTS	214	193	194	215	188	222	0	1226
SemiCTS (Ours)	101	82	82	100	80	100	4360	4905

Unlab.: Unlabeled images; Total: Total number of images.

3.2. Proposed CTSMatch Framework

We evaluated three baseline models on the SemiCTS dataset: (1) Dense-TL-Aug, the top-performing fully supervised model from prior CTS classification work [14]; (2) FixMatch; and (3) FreeMatch. We then introduce CTSMatch with its adaptive pseudo-label weighting mechanism.

3.2.1. Baseline Model Implementation

We benchmarked three models on the SemiCTS dataset. The first is Dense-TL-Aug, the top-performing fully supervised model from prior CTS classification work [14]; the other two are FixMatch and FreeMatch, two well-established semi-supervised learning methods.

Dense-TL-Aug combines a DenseNet backbone with transfer learning and automatic data augmentation to reduce overfitting, achieving 94.15% accuracy on the original CTS dataset. As a fully supervised method, it cannot use unlabeled data, which limits performance on SemiCTS when labeled data are scarce.

FixMatch [34] uses pseudo-labeling and consistency regularization with a high, fixed confidence threshold (typically 0.9 or 0.95) to filter pseudo-labels from unlabeled data. This ensures high-quality pseudo-labels but discards potentially correct low-confidence predictions, which hinders learning of diverse semantic features, particularly for datasets with weak-style characteristics.

FreeMatch [36] improves FixMatch by using adaptive, category-specific confidence thresholds that evolve during training, starting lower and increasing over time. This allows the model to learn from more pseudo-labels, especially during early training phases.

Benchmark goal: We evaluate whether semi-supervised methods can outperform a fully supervised baseline by using the abundant unlabeled data in SemiCTS, particularly when trained with limited labeled samples.

3.2.2. The Proposed CTSMatch Framework

The proposed CTSMatch framework integrates two core modules: SoftMatch for adaptive pseudo-label weighting and transfer learning for robust feature extraction. Figure 3 illustrates the overall architecture.

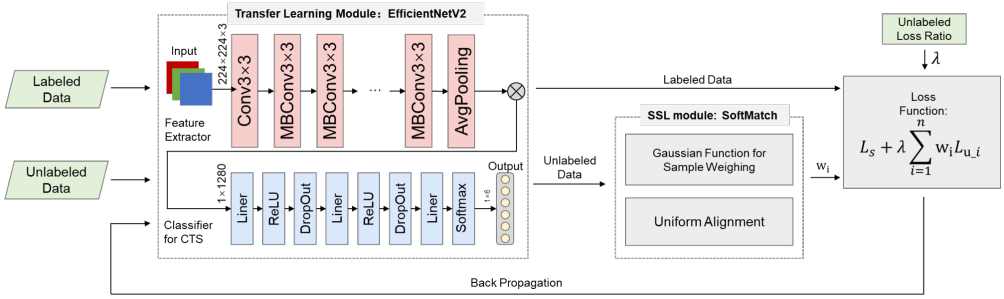

Figure 3. CTSMatch framework architecture. The system integrates two core modules: (1) a transfer learning module with EfficientNetV2 pretrained on ImageNet, and (2) a SoftMatch-based semi-supervised module that adaptively weights pseudo-labels according to confidence distribution. The total loss combines supervised loss (L_s) and weighted unsupervised loss (L_u), balanced by hyperparameter λ.

(1) Semi-supervised learning with SoftMatch. We selected SoftMatch as the semi-supervised core for two reasons. First, it performs well on image-classification benchmarks. Second, and more important for CTS data, it weights pseudo-labels continuously instead of filtering them with a hard threshold, which helps when stylistic cues are incomplete or ambiguous. During training, pseudo-labels from unlabeled images are weighted according to prediction confidence, modeled with a truncated Gaussian distribution updated by exponential moving average (EMA). For each batch, we estimate the mean and variance of pseudo-label confidence online, and the resulting weights modulate the cross-entropy loss for unlabeled samples. Algorithm 1 summarizes this process.

(2) Transfer learning with EfficientNetV2. We apply transfer learning [42] to improve generalization under sparse supervision. A backbone pretrained on ImageNet provides generic visual features (texture, shape, and contextual structure) that transfer to CTS classification. During training, we freeze the pretrained layers to preserve these low-level representations and update only the newly added classifier layers.

Algorithm 1 Adaptive pseudo-label weighting in SoftMatch.

Require: Batch size B; model predictions $\{(p_i, \Omega_i)\}_{i=1}^{B}$, where p_i is the predicted probability distribution after weak augmentation and Ω_i is the pseudo-label derived from strong augmentation; previous EMA statistics $(\mu_{t-1}, \sigma_{t-1}^2)$; supervised loss L_s; unsupervised loss weight λ; EMA momentum m.

Initialize: $\mu_b \leftarrow 0, \quad \sigma_b^2 \leftarrow 0, \quad L_u \leftarrow 0$

Step 1: Compute batch statistics on maximum prediction probability.

$$\mu_b \leftarrow \frac{1}{B}\sum_{i=1}^{B}\max(p_i)$$

$$\sigma_b^2 \leftarrow \frac{1}{B}\sum_{i=1}^{B}\big(\max(p_i) - \mu_b\big)^2$$

Step 2: Update Exponential Moving Average (EMA) statistics.

$\mu_t \leftarrow m\,\mu_{t-1} + (1-m)\,\mu_b$ ▷ EMA mean

$\sigma_t^2 \leftarrow m\,\sigma_{t-1}^2 + (1-m)\,\sigma_b^2$ ▷ EMA variance

Step 3: Compute weighted unsupervised loss.

Algorithm 1 *Cont.*

for $i = 1$ **to** B **do**
  if $\max(p_i) < \mu_t$ **then** ▷ attenuate low-confidence samples via truncated Gaussian
    $w_i \leftarrow \exp\left(-\frac{(\max(p_i) - \mu_t)^2}{2\sigma_t^2}\right)$
  else
    $w_i \leftarrow 1$ ▷ high-confidence samples contribute fully
  end if
  $L_u \leftarrow L_u + w_i \mathcal{H}(p_i, \Omega_i)$ ▷ weighted cross-entropy
end for

Ensure: $L = L_s + \lambda \frac{L_u}{B}$

We selected EfficientNetV2_L [43] as the backbone because its compound-scaling strategy balances depth, width, and resolution efficiently. Its strong ImageNet performance and feature-fusion design make it well suited to weak-style CTS imagery, where discriminative cues may be subtle and distributed across multiple architectural elements. Its stochastic-depth mechanism also improves robustness when pseudo-labels are noisy.

To better capture weak-style characteristics in CTS imagery, we add a hierarchical classifier to the backbone:

`Linear(1024)` → `ReLU` → `Dropout(0.5)` → `Linear(512)` → `ReLU` → `Dropout(0.3)` → `Linear(6)` → `Softmax`

Compared with a single fully connected layer, this design supports progressive feature composition and reduces overfitting.

3.3. Evaluation Metrics

We evaluate model performance using confusion matrices, which summarize classification results as True Positives (TPs), False Positives (FPs), False Negatives (FNs), and True Negatives (TNs) for each class. From these values we compute four standard metrics, defined in Equations (1)–(4): *Accuracy*, *Precision*, *Recall*, and *F1 score.*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision} = \frac{2TP}{2TP + FP + FN} \tag{4}$$

To account for class imbalances, we calculate these metrics for each architectural style category. We report macro-averaged values across all categories so each class contributes equally to the overall performance assessment.

3.4. Style Lineage Analysis

In addition to evaluating CTSMatch classification performance, we conduct a post hoc feature-space analysis to examine intra-style heterogeneity, hierarchical relationships among styles, and outlier cases. Because this interpretive analysis is intended to characterize stylistic structure rather than to modify the CTSMatch classifier, it proceeds in four stages.

(1) Deep feature extraction. We use ResNet50 [44], which addresses vanishing gradients and degradation in very deep networks through residual connections, to extract deep features:

$$\mathbf{f}_i = \phi(\mathbf{x}_i) \in \mathbb{R}^d, \tag{5}$$

where ϕ denotes the ResNet50 penultimate-layer extractor, $\mathbf{x}_i$ is the i-th input image, and $d = 2048$ is the feature dimensionality.

(2) Dimensionality reduction. We apply Principal Component Analysis (PCA) to project the original features onto orthogonal principal components that maximize explained variance:

$$\mathbf{z}_i = \mathbf{W}_{\text{PCA}}^{\top} \mathbf{f}_i \in \mathbb{R}^K, \tag{6}$$

where $\mathbf{W}_{\text{PCA}} \in \mathbb{R}^{d \times K}$ is the PCA projection matrix retaining the first $K = 50$ principal components.

(3) Quantitative intra-style heterogeneity analysis. We quantify heterogeneity with three complementary metrics capturing central tendency, boundary ambiguity, and local sparsity within each style cluster: z-scores [45], neighbor consistency based on k-nearest neighbors [46], and a local density variation coefficient [47,48].

Z-score measures how far a sample deviates from the style prototype:

$$z_i = \frac{\|\mathbf{z}_i - \boldsymbol{\mu}_c\|}{\sigma_c} \tag{7}$$

where $\boldsymbol{\mu}_c = \frac{1}{n_c} \sum_{j \in C_c} \mathbf{z}_j$ is the centroid of cluster c, and σ_c is the standard deviation within the cluster. Samples with $z_i > 2$ are flagged as weak styles.

Nearest neighbor consistency measures boundary proximity. Samples in homogeneous regions have high consistency, whereas boundary samples have low consistency:

$$\text{consistency}_i = \frac{1}{k} \sum_{j \in \mathcal{N}_k(\mathbf{z}_i)} \mathbb{I}(y_j = y_i) \tag{8}$$

where $\mathcal{N}_k(\mathbf{z}_i)$ denotes the set of k nearest neighbors, and $\mathbb{I}(\cdot)$ is the indicator function. We use $k = 10$ in our implementation.

Local density variation coefficient captures local structure density based on the inverse mean distance to the k nearest neighbors:

$$\rho_i = \frac{1}{\frac{1}{k} \sum_{j \in \mathcal{N}_k(\mathbf{z}_i)} \|\mathbf{z}_i - \mathbf{z}_j\| + \epsilon} \tag{9}$$

where $\bar{d}_k(\mathbf{z}_i)$ is the mean distance to the k nearest neighbors and $\epsilon = 10^{-6}$ is a small constant that prevents division by zero.

(4) Mapping style transitions and detecting outliers. Architectural styles form a continuum shaped by cultural diffusion and geographic proximity, not isolated categories. We use t-Distributed Stochastic Neighbor Embedding (t-SNE) [49] to project the high-dimensional feature space into two dimensions, exposing transition zones between styles. We first compute pairwise similarities in the high-dimensional space:

$$p_{j|i} = \frac{\exp(-\|\mathbf{z}_i - \mathbf{z}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{z}_i - \mathbf{z}_k\|^2 / 2\sigma_i^2)} \tag{10}$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n} \tag{11}$$

We then compute pairwise similarities in the low-dimensional space using Student's t-distribution, whose heavier tails preserve local neighborhoods:

$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}} \tag{12}$$

The low-dimensional embedding minimizes the Kullback–Leibler divergence between the two similarity distributions:

$$\mathcal{L} = \mathrm{KL}(P\|Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{13}$$

This visualization reveals broader relationships among styles, including overlap, transition zones, and cluster boundaries. Samples far from their assigned style center or in ambiguous positions between clusters may represent regional variants, hybrid cases, mislabeled instances, or novel typologies.

Accuracy measures how well the model classifies styles, whereas feature-space analysis explains how styles organize internally and relate to one another. This combination advances both label-efficient classification and interpretation of stylistic variation in traditional Chinese architecture, supporting geospatial analysis of cultural heritage patterns.

4. Experimental Results

All experiments were conducted on SemiCTS, as described in Section 3.1. We used Stochastic Gradient Descent (SGD) with momentum 0.9, an initial learning rate of 0.03, and cosine annealing. Training ran for 1600 iterations with a batch size of 64. For the semi-supervised models (FixMatch, FreeMatch, and CTSMatch), we applied strong augmentation seven times per batch to the unlabeled data. The unsupervised loss weight λ was set to 0.3. To improve robustness in the single-run setting,we fixed the random seed (42) for all models and used a deterministic stratified village-level split of the labeled data (70% training, 30% validation) rather than per-image random sampling. This design prevents data leakage, because images from the same village never appear in both training and validation sets, and it also reduces sampling variance.

4.1. Baseline Performance

We compare CTSMatch against three baselines: Dense-TL-Aug (a fully supervised transfer learning model) and two semi-supervised methods (FixMatch and FreeMatch). As shown in Table 3 and Figure 4, Dense-TL-Aug achieves 94.15% accuracy, whereas FixMatch and FreeMatch both reach 93.87%. Standard semi-supervised methods do not improve on the fully supervised baseline on SemiCTS. This shortfall reflects the weak-style phenomenon: when multi-view architectural samples exhibit high intra-class variation, threshold-based pseudo-label filtering discards many correct predictions and therefore weakens, rather than enhances, learning. The learning curves in Figure 4 support this interpretation: the SSL baselines converge slowly because conservative thresholds limit pseudo-label availability, whereas Dense-TL-Aug oscillates under aggressive multi-view augmentation.

Table 3. Performance comparison of models on the SemiCTS dataset.

Model	Accuracy	Precision	Recall	F1 Score
Dense-TL-Aug	94.15%	96.53%	94.15%	95.06%
FixMatch	93.87%	93.73%	93.87%	93.67%
FreeMatch	93.87%	93.77%	94.00%	93.76%
CTSMatch (Ours)	**96.93%**	**96.78%**	**96.78%**	**96.75%**

Note: Results are from a single representative run with fixed random seed and deterministic data split. The smooth convergence (Figure 5) and consistent performance across validation checkpoints suggest high stability.

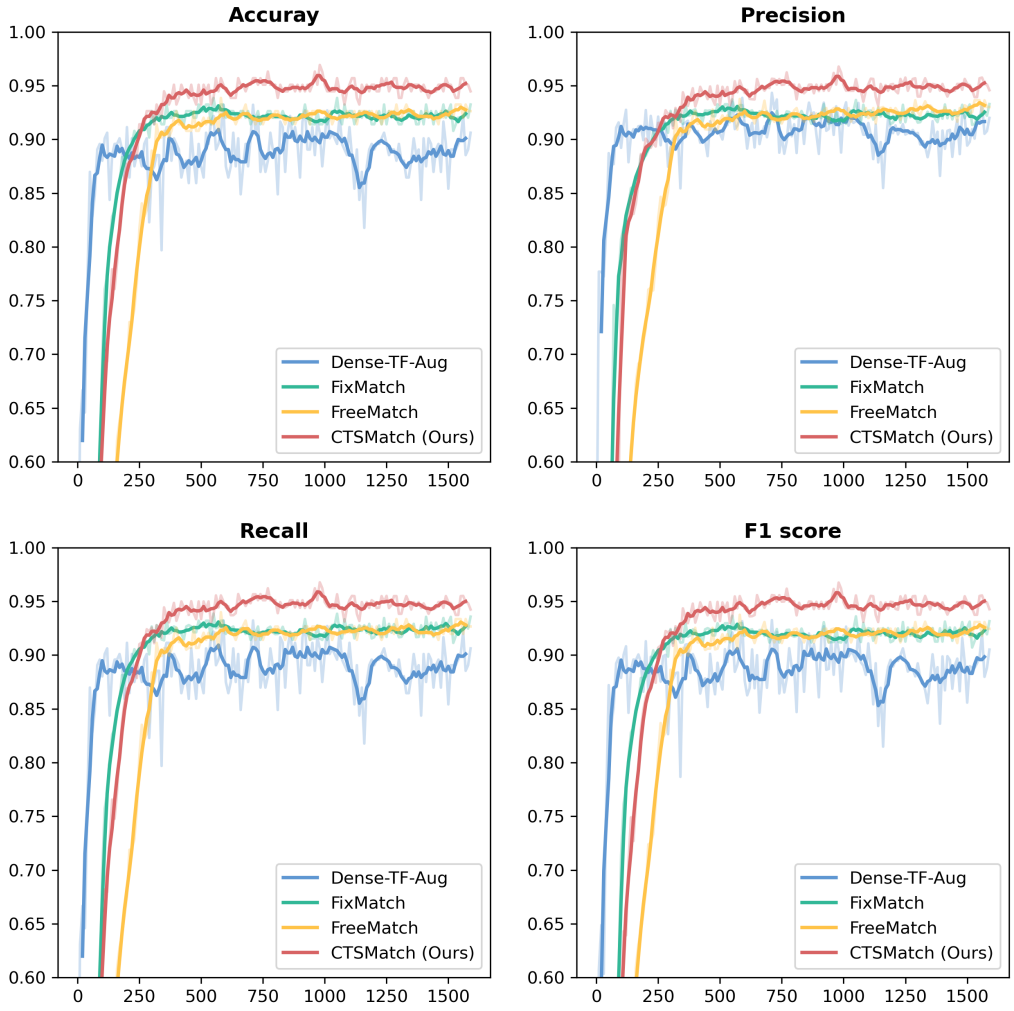

Figure 4. Learning curves showing Accuracy, Precision, Recall, and F1 score during training for Dense-TL-Aug, FixMatch, FreeMatch, and CTSMatch (ours).

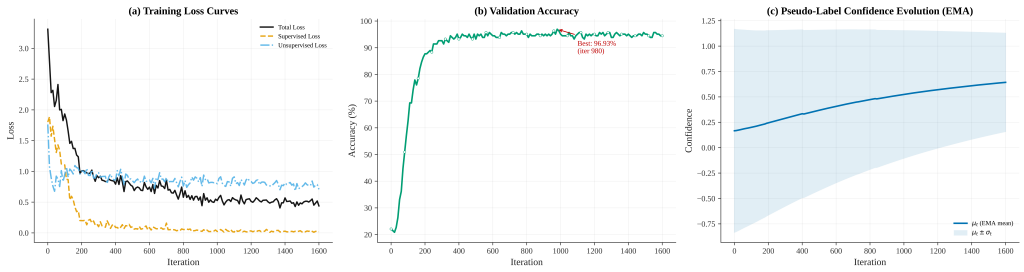

Figure 5. Detailed convergence diagnostics for CTSMatch. The panels show total/supervised/unsupervised losses, validation accuracy and F1, EMA pseudo-label confidence with variance band across 1600 training iterations.

4.2. Performance of the Proposed CTSMatch Framework

CTSMatch addresses this limitation by replacing hard pseudo-label filtering with adaptive confidence-weighted pseudo-labels. Rather than discarding low-confidence predictions, CTSMatch down-weights them according to their position in the confidence distribution, thereby preserving information from weak-style samples. On SemiCTS, CTSMatch achieves 96.93% accuracy and outperforms all baselines. Relative to FixMatch and FreeMatch, CTSMatch improves accuracy by 3.06 percentage points. More importantly, it exceeds the 94.15% accuracy of Dense-TL-Aug [14] while using only 545 labeled images, which is approximately 40% fewer than the 1226 labeled samples used by the fully super-

vised approach. The aggregate learning curves in Figure 4 and the detailed diagnostics in Figure 5 show that adaptive weighting enables more effective use of unlabeled data and leads to stable optimization. CTSMatch reaches 90%, 95%, and 98% of its best validation accuracy by iterations 200, 280, and 390, respectively, and attains its peak accuracy of 96.93% at iteration 980.

4.3. Additional Diagnostic Evaluation

Training convergence behavior. Because the semi-supervised implementation is iteration-based rather than epoch-based, we report convergence in training iterations. Figure 5 shows the complete CTSMatch dynamics over 1600 iterations. The total, supervised, and unsupervised losses decrease smoothly without divergence, while the EMA pseudo-label confidence increases from 0.20 at initialization to 0.73 at the end of training. This indicates that performance gains arise from stable optimization rather than from isolated late-stage checkpoints.

Computational complexity and inference time. Figure 6 compares the efficiency of the evaluated backbones. DenseNet121-based Dense-TL-Aug, FixMatch, and FreeMatch each require 12.2 M total parameters, 5.80 GFLOPs per image, and 16.2 ms per image (61.8 img/s) on a Tesla T4. CTSMatch uses an EfficientNetV2-L backbone and therefore requires 123.5 M total parameters, 24.77 GFLOPs, and 37.6 ms per image (26.6 img/s). However, because the pretrained backbone remains frozen during training, the trainable parameter count stays comparable to the DenseNet121 baselines retained (6.3 M versus 5.2 M). The added computational cost is therefore a practical accuracy–efficiency trade-off rather than a substantially larger optimization burden.

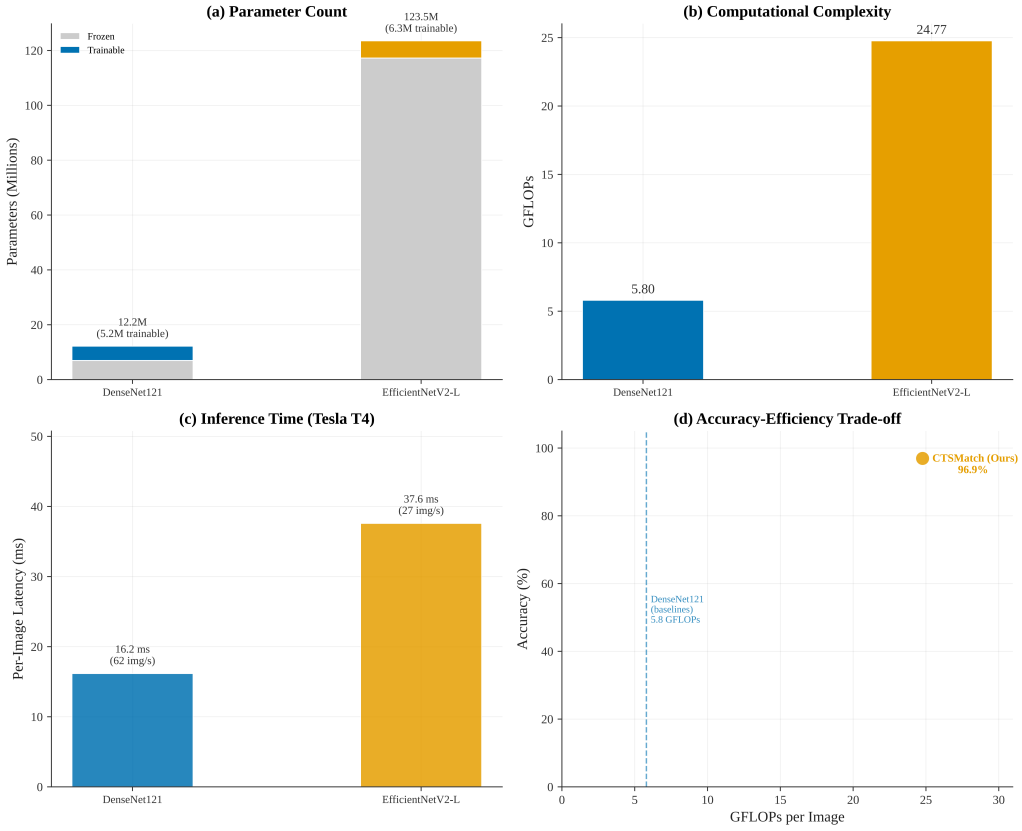

Figure 6. Model complexity and inference-efficiency comparison for Dense-TL-Aug, FixMatch, FreeMatch, and CTSMatch. The figure summarizes total parameters, trainable parameters, GFLOPs, inference time, and the resulting accuracy–efficiency trade-off.

Pseudo-label reliability and error propagation. Figure 7 examines whether inaccurate pseudo-labels propagate through training. Errors are concentrated in low- and medium-confidence samples that SoftMatch already down-weights: the error rate is 4.7% for predictions below 0.70 confidence, 2.7% for predictions between 0.70 and 0.95, and 0% for the high-confidence bin ($\geq$ 0.95). Correct predictions also show substantially higher mean confidence (0.741 vs. 0.555) and larger class-margin separation (0.640 vs. 0.398) than incorrect predictions. Only 5 of 163 validation samples are misclassified, and these errors are concentrated in the transitional styles Jing, Su, and Wan. The expected calibration error is 0.234, indicating slightly conservative calibration rather than overconfident error amplification. Together, these diagnostics show that adaptive weighting suppresses ambiguous pseudo-labels before they can dominate training.

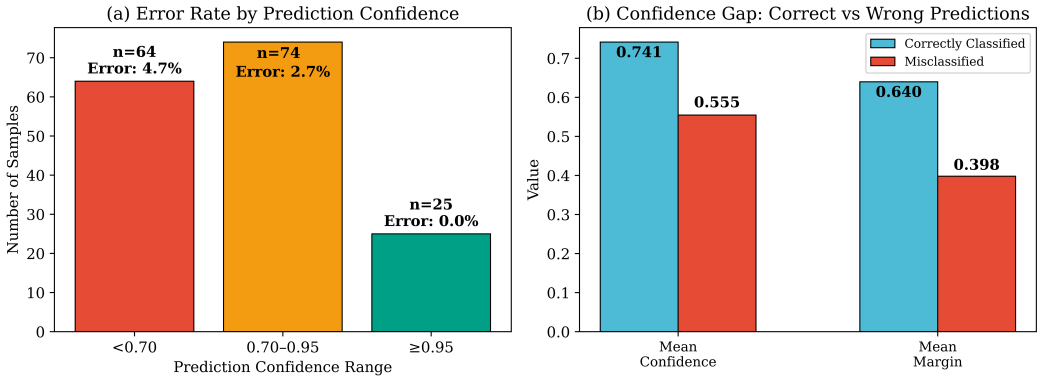

Figure 7. Pseudo-label reliability analysis for CTSMatch. The panels summarize confidence-stratified error rates, reliability calibration, confidence-gap statistics, and the evolution of the pseudo-label quality proxy during training.

5. Architectural Style Lineage Analysis

CTSMatch achieves high accuracy in separating six predefined architectural styles, but accuracy alone cannot explain how styles organize in feature space or where difficult boundary cases emerge. For cultural heritage research, understanding how and why styles differ matters as much as classification. We examine the learned feature space from three perspectives: intra-style heterogeneity, inter-style similarity and transition, and outlier cases. Together, these analyses turn the classification results into an interpretable basis for cultural heritage discovery, revealing spatial differentiation and stylistic lineage patterns that inform geospatial documentation and conservation planning.

5.1. Intra-Style Heterogeneity and the Weak-Style Phenomenon

To evaluate the weak-style phenomenon within the six style categories, we extract ResNet50 features from the penultimate layer and reduce dimensionality to 50 components via PCA. We then characterize semantic variation using three metrics: z-score (standardized distance from each sample to its style center), neighbor consistency (proportion of $k = 10$ nearest neighbors sharing the same label), and local density (inverse of mean distance to neighbors). As Figure 8 shows, all six styles exhibit measurable heterogeneity. Weak-style effects represent a general property of multi-view architectural imagery rather than an anomaly in a single category.

Absolute differences among styles are modest across metrics, with neighbor consistency dominating the combined heterogeneity score. Jin is the most compact style (highest neighbor consistency: 0.38), suggesting that its samples cluster around a relatively stable visual prototype. This compactness aligns with the geographic concentration of Jin archi-

tecture in the Loess Plateau, where regional isolation limited stylistic mixing. By contrast, Su, Wan, and Min exhibit stronger internal variation, each for different reasons.

Su and Wan exhibit the highest z-score, which indicates that their samples are widely dispersed in feature space and do not converge on a single visual prototype. This dispersion reflects the geographical position of Su and Wan style architectures, where architectural exchange with neighboring traditions is common; for example, Suzhou gardens, water-town dwellings, and ancestral halls across the Taihu region. As a result, the Su and Wan styles manifest a radiative heterogeneity rather than a unified core.

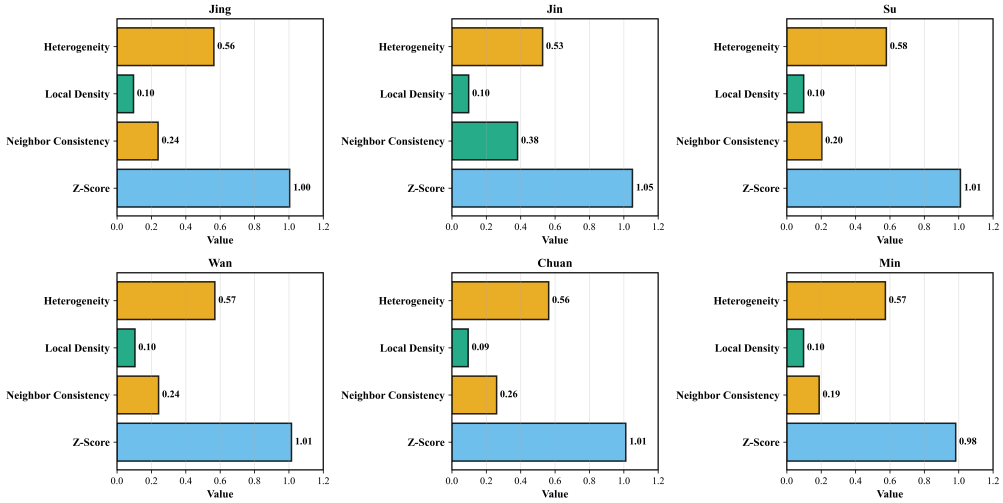

Figure 8. Intra-style heterogeneity metrics across six architectural traditions. For each style, the panel shows the combined heterogeneity score, z-score (standardized distance from style center), neighbor consistency (proportion of $k = 10$ nearest neighbors sharing the same label), and local density (inverse of mean distance to neighbors). Higher values indicate greater heterogeneity or dispersion within the style cluster.

Min exhibits the highest local density variation, displaying a clear core–periphery structure in feature space. Iconic building types—primarily Tulou communal fortresses and red-brick dwellings—form a dense central cluster, while local variants distribute more diffusely around this nucleus. Fujian's mountainous terrain, which historically isolated sub-regions from one another, explains this pattern: distinctive local architectures evolved in relative isolation, creating a dense core of widely recognized forms surrounded by a periphery of regional variants.

The feature space reveals that Chinese traditional architectural styles are not sharply separated categories. Each tradition shows a distinct heterogeneity pattern shaped by geography, historical exchange, and viewpoint variation. These findings provide the foundation for the broader lineage analysis that follows.

5.2. Stylistic Continuum and Hierarchical Lineage

The heterogeneity analysis raises a broader question: are Chinese traditional architectural styles best understood as discrete classes, or as a continuum with substantial overlap and hybridization? To reduce subjectivity in this interpretation, we computed silhouette coefficients in three feature spaces. The overall silhouette score is 0.0183 in the raw 2048D embedding, 0.0358 after PCA to 50 dimensions, and 0.0249 in the 2D projection. These low but positive values indicate weak yet non-random cluster separation: the six styles are morphologically structured in feature space, but they are not perfectly isolated categories. At the class level, Wan (0.0607) and Chuan (0.0591) show the strongest cohesion, Jin (0.0347) and Min (0.0344) occupy an intermediate position, and Jing (0.0034) and Su (0.0085) are

the most diffuse. Combined with neighbor consistency and centroid-level cosine similarity, this quantitative evidence supports our interpretation of the six styles as overlapping but still meaningful morphological entities. Figure 9 summarizes these separability diagnostics, and Figure 10 shows the corresponding inter-style relationships.

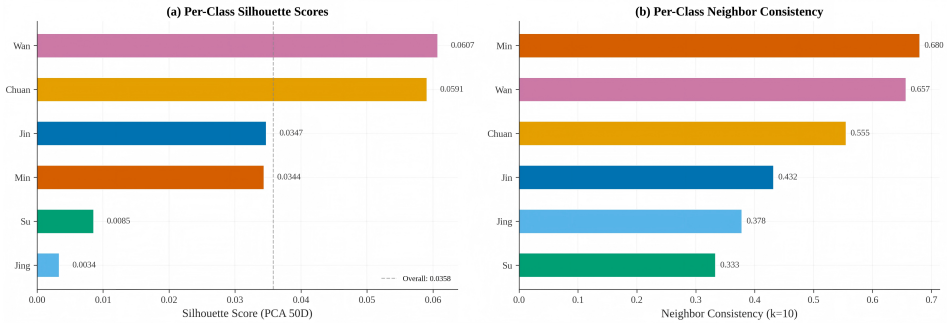

Figure 9. Quantitative style-separability diagnostics. The panels summarize per-class silhouette coefficients and neighbor consistency.

Inter-style similarity patterns. The visualization in Figure 10 shows patterns of inter-style similarity. The most similar style pairs are Min–Chuan (0.438), Wan–Min (0.210), Jin–Su (0.189), and Jing–Jin (0.130), reflecting geographic adjacency. The most dissimilar pairs are Jin–Chuan (−0.743), Jing–Min (−0.622), and Jing–Wan (−0.564). These negative correlations align with the architectural dichotomy between northern traditions (heavy, symmetrical, earthen construction) and southern traditions (light, flexible, mountain-adapted forms). Panel (b) shows that Chuan exhibits the most hybrid samples, lying at substantial distances from their assigned style center, with hybrid transitions occurring primarily with Min. These findings support a continuum view: regional clusters exist (Min–Chuan–Wan, Jing–Jin–Su), but the overall architectural landscape is better characterized as overlapping gradients than as discrete categories.

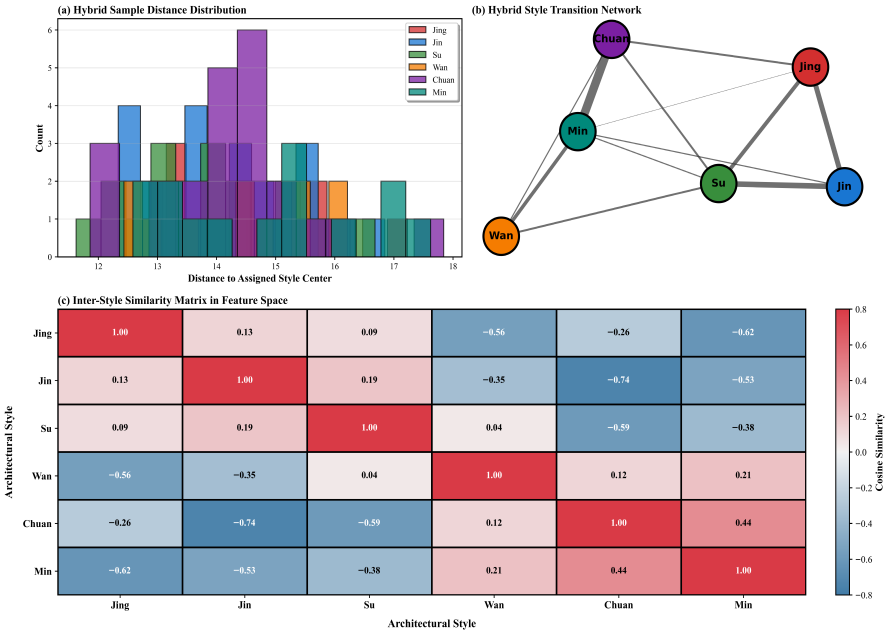

Figure 10. Stylistic continuum analysis revealing inter-style relationships. (**a**) Hybrid samples plotted by distance from their assigned style center. (**b**) Network graph showing hybrid connections between styles, where edge thickness represents the number of hybrid samples. (**c**) Pairwise cosine similarity heatmap, where darker blue indicates greater similarity and lighter orange indicates greater dissimilarity. The results support a continuum view rather than discrete style categories.

Hierarchical clustering. The dendrogram in Figure 11 confirms the similarity matrix, showing two clusters: a northern group (Jing, Jin, Su) and a southern group (Chuan, Min, Wan). Su connects elements from both regions, serving as an intermediate style between the two clusters. Branch lengths indicate stronger similarity within regional clusters than across them, supporting the view that geographic proximity and historical exchange shape the style continuum.

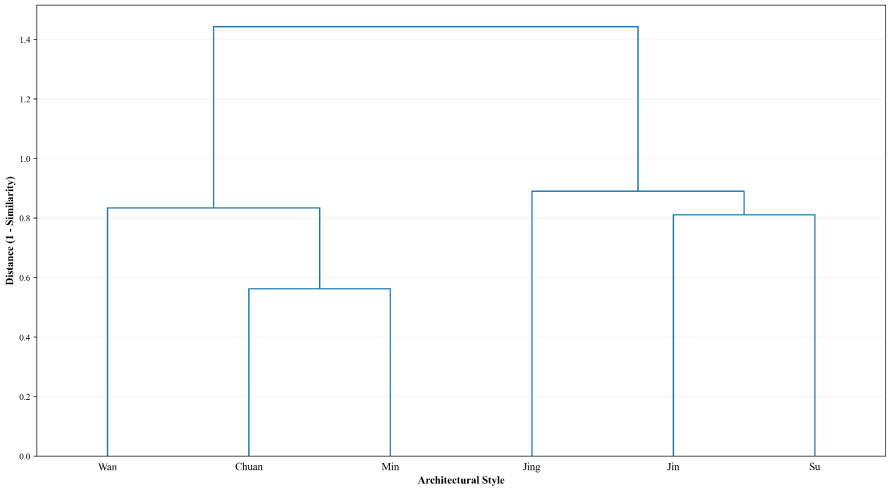

Figure 11. Hierarchical clustering dendrogram of architectural styles. The tree with blue branches reveals two regional groupings: a northern cluster (Jing, Jin, Su) and a southern cluster (Chuan, Min, Wan), with Su serving as an intermediate bridge. Shorter branch lengths indicate greater similarity between styles.

5.3. Outliers, Ambiguous Cases, and Novel Cases

The two-dimensional t-SNE projection in Figure 12 shows substantial overlap among the architectural style clusters. Rather than forming clearly separated regions, the styles exhibit broad boundary zones in which samples cannot be assigned unambiguously. The relative positions of cluster centers are consistent with the similarity analysis: Jin–Su and Jing–Jin lie close to one another, whereas Jin–Chuan and Jing–Min are farthest apart. This spatial arrangement further supports the continuum hypothesis, suggesting that architectural styles in Chinese traditional settlements are better understood as overlapping gradients than as discrete categories. The presence of these boundary cases does not indicate model failure. Instead, it reflects the inherent continuity of architectural traditions, in which regional variants and hybrids are genuine cultural phenomena rather than classification errors.

Outlier statistics and geographic patterns are shown in Figure 13. Quantitative outlier analysis clarifies boundary uncertainty. Among all samples, 15% exhibit distinctive architectural forms: 8.7% are regional variants (adaptations to local conditions) and 6.3% are hybrid styles (blended characteristics from multiple traditions). An additional 34.3% of samples fall into low-confidence classifications. Outlier rates vary across styles: Chuan shows the highest outlier rate (23.1%), while Jin shows the lowest (10.1%). Geographically, this pattern aligns with regional diversity: northern styles (Jin, Wan) show relative homogeneity, southern styles (Min, Chuan) display greater diversity, and central styles (Jing, Su) occupy intermediate positions.

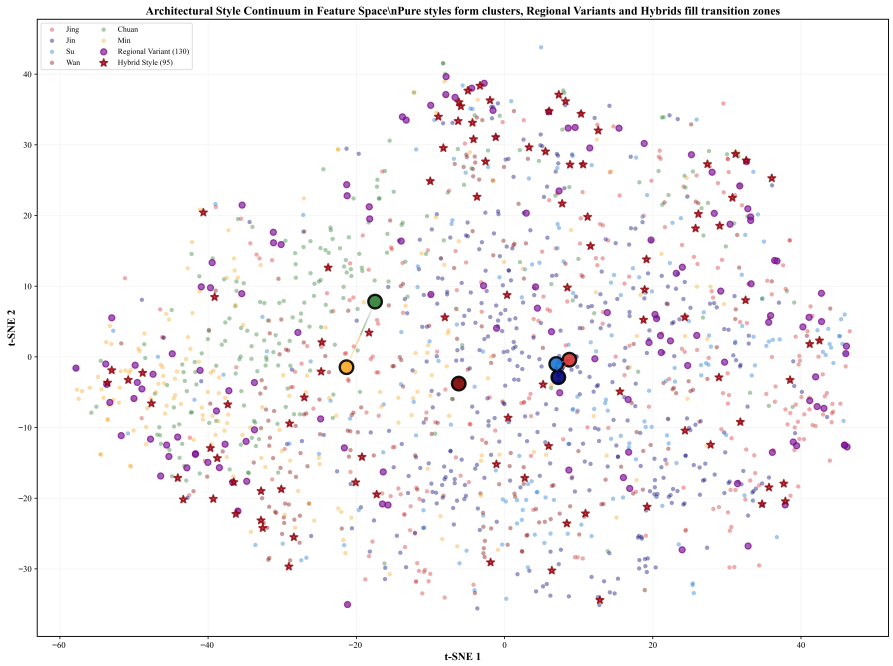

Figure 12. Two-dimensional t-SNE visualization of the architectural style feature space. The overlapping cluster boundaries and numerous boundary cases support the stylistic continuum hypothesis. Colors indicate the six predefined architectural styles, with overlapping regions revealing transitional zones and samples that defy discrete classification.

Regional variants and hybrid styles. Figure 14 shows the top 20 outliers among 225 detected samples. Regional variants receive relatively confident predictions while occupying peripheral positions in their style clusters, indicating that the model captures them as legitimate members while preserving local distinctiveness. These variants arise from: (1) geographic isolation leading to material and terrain adaptation (sample No. 1 is a Chuan-style structure with an unusual roof form associated with bridge architecture); (2) functional specialization (sample No. 5 was built as a representative village entrance rather than as a residence); or (3) unusual acquisition viewpoints.

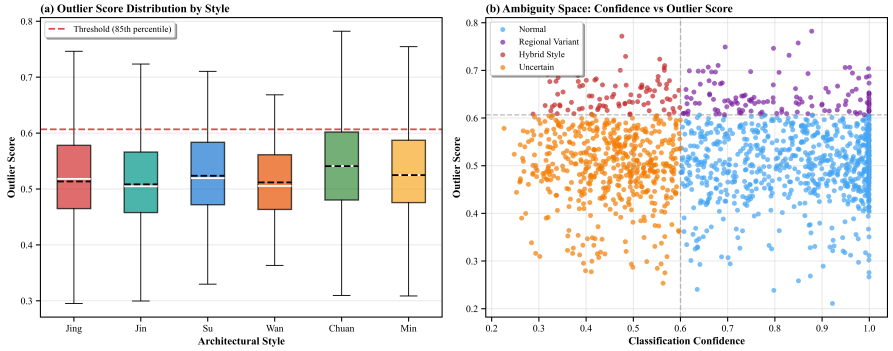

Figure 13. Outlier analysis results. (**a**) Box plots showing outlier score distribution for each architectural style, with Chuan exhibiting the highest variability. (**b**) Ambiguity space scatter plot defining outlier cases along two dimensions: classification confidence (x-axis) and outlier score (y-axis). Low-confidence, high-outlier samples represent the most ambiguous cases.

Hybrid styles exhibit blended characteristics from multiple architectural traditions. These samples show core features of two or more styles and occupy transition zones between clusters in feature space, making unambiguous assignment to any single category difficult. The formation of hybrid styles reflects four mechanisms: (1) cultural diffusion and inter-regional exchange along trade routes and migration corridors; (2) emergence of

geographical transition zones where architectural spheres overlap; (3) superimposition of official (imperial or monastic) architectural paradigms onto vernacular traditions; and (4) accretion of stylistic elements across historical periods. A small subset of samples exhibits novel characteristics that do not align with any of the six predefined styles, representing cases for further investigation through field study or archival research.

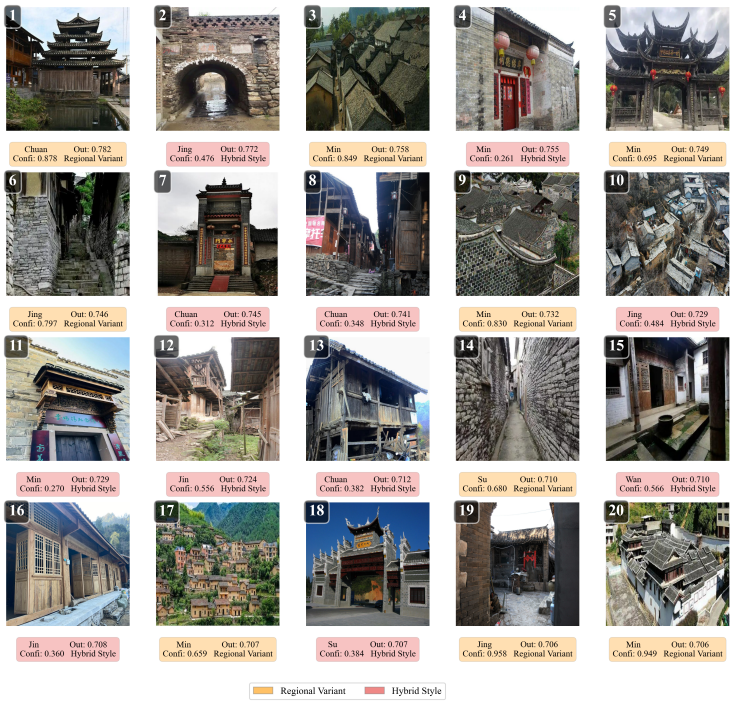

Figure 14. Top 20 outlier samples ranked by outlier score. These architectural images exhibit the highest deviation from their assigned style clusters in feature space. The grid reveals two categories of outliers: regional variants (adaptations to local geography or materials) and hybrid styles (samples blending characteristics from multiple architectural traditions). Each sample is annotated with its rank number.

6. Discussion

This study addresses a core bottleneck in classifying architectural styles of Chinese traditional settlements: how to learn reliable style distinctions when expert labels are scarce and architectural cues are weakly separated. Using only 545 labeled images, CTSMatch achieves 96.93% accuracy on SemiCTS and outperforms both the strongest fully supervised baseline and conventional semi-supervised baselines. This result indicates that adaptive confidence weighting is better suited to weak-style CTS imagery than threshold-based pseudo-label filtering. More broadly, the findings suggest that semi-supervised learning can serve as a practical strategy for heritage documentation when expert annotation is costly and difficult to scale.

Practical trade-offs and convergence behavior: CTSMatch reaches 98% of its best validation accuracy by iteration 390 and peaks at iteration 980, indicating stable optimization within the current single-split experimental setting. Although inference is slower than that of the DenseNet121-based baselines (37.6 ms/image vs. 16.2 ms/image), the number of trainable parameters remains comparable because the EfficientNetV2-L backbone is frozen. The performance gain should therefore be interpreted as a practical accuracy–efficiency trade-off rather than as the consequence of a substantially larger trainable model.

Error propagation under adaptive weighting: Validation errors concentrate in low- and medium-confidence pseudo-label regions, with no errors in the high-confidence bin ($\geq$0.95). Incorrect predictions exhibit lower confidence and smaller class margins than

correct ones, and the remaining errors are concentrated in the transitional styles Jing, Su, and Wan, which also appear as the most diffuse categories in the feature-space analysis. This pattern suggests that SoftMatch's adaptive weighting limits error propagation by assigning conservative weights to ambiguous samples rather than reinforcing them, which is particularly important in CTS imagery where blurred stylistic boundaries are part of the domain rather than an anomaly.

Stylistic continuum and quantitative support: Feature-space analysis reveals that the six architectural traditions are structured but only weakly separated in embedding space. Jin exhibits the most compact feature distribution, whereas Su, Wan, and Min show much greater internal dispersion. The learned embeddings suggest a stylistic continuum in which northern styles (Jing, Jin, Su) and southern styles (Chuan, Min, Wan) form regional groupings but remain connected by gradual transitions. In total, 15% of the samples appear as distinctive regional variants (8.7%) or hybrid forms (6.3%), and a further 34.3% fall into low-confidence boundary regions. This interpretation is supported by silhouette analysis: the scores in the raw, PCA50-reduced, and 2D t-SNE spaces remain close to zero but positive, which is consistent with weak but non-random separation and therefore favors a continuum interpretation over sharply discrete clusters.

Limitations: The current results suggest that the combination of an ImageNet-pretrained backbone, SoftMatch's down-weighting of low-confidence samples, and explicit outlier analysis may help reduce the influence of contemporary visual noise, such as modern infrastructure or vegetation occlusion; however, this effect was not directly evaluated in a controlled experiment. Two substantive limitations therefore remain. First, our closed-set formulation cannot identify novel styles beyond the six predefined categories. Second, pixel-level visual features may not fully capture the component-level semantics, such as dougong bracket forms and ridge decorations, that architectural experts use in practice. As a complement to broader geomatics-oriented heritage studies [48], the present work focuses on label-efficient ground-image classification while providing an interpretable basis for subsequent lineage and spatial analysis.

7. Conclusions

This study addresses a practical bottleneck in architectural heritage documentation: how to classify Chinese traditional settlement styles when expert labels are scarce and stylistic boundaries are weak. Using only 545 labeled images, CTSMatch achieves 96.93% accuracy on SemiCTS and shows that semi-supervised learning can improve classification performance under limited annotation. The resulting analyses further indicate that Chinese traditional architecture is better understood as a structured but overlapping stylistic continuum rather than a set of sharply separated categories. CTSMatch therefore contributes both a practical label-efficient classification framework and an analytical basis for studying stylistic differentiation, transition, and atypical cases in Chinese traditional settlements.

Future work should proceed in three directions. First, open-set recognition and hierarchical labeling are needed to identify novel or mixed architectural styles beyond the current six-class closed set. Second, attention- or segmentation-based preprocessing should be explored to suppress visual noise and better align machine perception with component-level architectural semantics. Third, the release of SemiCTS geographic metadata should support future fusion with satellite-derived environmental layers and broader spatial validation. The SemiCTS dataset and CTSMatch framework are publicly available at https://github.com/wzc32/CTSMatch (accessed on 27 March 2026).

Author Contributions: Conceptualization, Qing Han and Chao Yin; methodology, Chao Yin and Zicheng Wang; software, Zicheng Wang; validation, Qing Han, Zicheng Wang, and Chao Yin; formal analysis, Zicheng Wang and Chao Yin; investigation, Zicheng Wang, Zhiwei Hou, and Tianci Yao; data curation, Zicheng Wang and Tianci Yao; writing—original draft preparation, Zicheng Wang and Qing Han; writing—review and editing, Chao Yin, Zhiwei Hou and Qing Han; supervision, Chao Yin; funding acquisition, Qing Han and Chao Yin. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Hunan Provincial Natural Science Foundation of China (No. 2023JJ40098), the China Postdoctoral Science Foundation (No. 2023M740761), the National Natural Science Foundation of China (No. 42001198), the National-Local Joint Engineering Laboratory on Digital Preservation and Innovative Technologies for the Culture of Traditional Villages and Towns (No. 2025HSKFJJ004 and 2025HSKFJJ009), the Science and Technology Program of Guangdong (NO.2024B1212080002).

Data Availability Statement: The SemiCTS dataset and CTSMatch framework developed during this study are publicly available on GitHub: https://github.com/wzc32/CTSMatch (accessed on 27 March 2026) . Further data supporting the findings are available within this article.

Acknowledgments: We thank the anonymous reviewers for their valuable feedback and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wei, X.; Si, Z. Fully Exploring Traditional Chinese Culture and Promoting Organic Development of Green City. *Procedia Eng.* **2017**, *180*, 1531–1540. [CrossRef]
2. Noble, A.G. *Traditional Buildings: A Global Survey of Structural Forms and Cultural Functions*; I.B. Tauris: London, UK, 2007.
3. Hamka; Winarni, S.Comparative Understanding of Traditional Architecture Based on Literature Review: Refining the Definition of Traditional Architecture. *Int. J. Archit. Urban.* **2021**, *5*, 252–262. [CrossRef]
4. Yavuz, K.B.; Savran, S. Glocalization and Tourism Relation: Bauhaus Architecture in Tel Aviv Example. *Idealkent* **2020**, *11*, 1205–1222. Available online: https://dergipark.org.tr/en/pub/idealkent/article/675530 (accessed on 14 May 2026).
5. Khamdevi, M.; Bott, H. Rethinking Tourism: Bali's Failure. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *126*, 012171. [CrossRef]
6. Fuentes, J.M. Methodological Bases for Documenting and Reusing Vernacular Farm Architecture. *J. Cult. Herit.* **2010**, *11*, 119–129. [CrossRef]
7. Chen, X. The Core of China's Rural Revitalization: Exerting the Functions of Rural Area. *China Agric. Econ. Rev.* **2020**, *12*, 1–13. [CrossRef]
8. Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C.W.; Choudhary, A.; Agrawal, A.; Billinge, S.J.L.; et al. Recent Advances and Applications of Deep Learning Methods in Materials Science. *npj Comput. Mater.* **2022**, *8*, 59. [CrossRef]
9. Kuncoro, I.R.A.; Setyaningsih, W.; Sunoko, K. Characteristics of architectural typology in colonial buildings in Loji Wetan Surakarta area. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *969*, 012063. [CrossRef]
10. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [CrossRef]
11. Kong, Y.; Xue, P.; Xu, Y.; Li, X. An Environmental Pattern Recognition Method for Traditional Chinese Settlements Using Deep Learning. *Appl. Sci.* **2023**, *13*, 4778. [CrossRef]
12. Xu, Z.; Tao, D.; Zhang, Y.; Wu, J.; Tsoi, A.C. Architectural Style Classification Using Multinomial Latent Logistic Regression. In *Computer Vision–ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; Volume 8689, pp. 600–615. [CrossRef]
13. Sun, M.; Zhang, F.; Duarte, F.; Ratti, C. Understanding Architecture Age and Style through Deep Learning. *Cities* **2022**, *128*, 103787. [CrossRef]
14. Han, Q.; Yin, C.; Deng, Y.; Liu, P. Towards Classification of Architectural Styles of Chinese Traditional Settlements Using Deep Learning: A Dataset, a New Framework, and Its Interpretability. *Remote Sens.* **2022**, *14*, 5250. [CrossRef]
15. Belhi, A.; Ahmed, H.O.; Alfaqheri, T.; Bouras, A.; Sadka, A.H.; Foufou, S. Study and Evaluation of Pre-trained CNN Networks for Cultural Heritage Image Classification. In *Data Analytics for Cultural Heritage*; Springer: Cham, Switzerland, 2021; pp. 47–69. [CrossRef]

16. Cosovic, M.; Jankovic, R.; Ramic-Brkic, B. Cultural Heritage Image Classification. In *Data Analytics for Cultural Heritage*; Springer: Cham, Switzerland, 2021; pp. 25–45. [CrossRef]
17. Zou, H.; Ge, J.; Liu, R.; He, L. Feature Recognition of Regional Architecture Forms Based on Machine Learning: A Case Study of Architecture Heritage in Hubei Province, China. *Sustainability* **2023**, *15*, 3504. [CrossRef]
18. Han, Q.; Yin, C.; Yang, Y. A Computational Approach to Traditional Chinese Architectural Styles: Intelligent Identification and Spatial Pattern Analysis. *Trop. Geogr.* **2026**, *46*, 179–189. (In Chinese) [CrossRef]
19. Rodríguez-Antuñano, I.; Sousa, J.J.; Bakoň, M.; Ruiz-Armenteros, A.M.; Martínez-Sánchez, J.; Riveiro, B. Empowering intermediate cities: Cost-effective heritage preservation through satellite remote sensing and deep learning. *Int. J. Remote Sens.* **2024**, *45*, 4046–4074. [CrossRef]
20. Chen, H.; Tao, R.; Fan, Y.; Wang, Y.; Wang, J.; Schiele, B.; Xie, X.; Raj, B.; Savvides, M. SoftMatch: Addressing the Quantity–Quality Trade-off in Semi-Supervised Learning. *arXiv* **2023**, arXiv:2301.10921. [CrossRef]
21. Hou, X.; Cheng, B.; Yang, J. A Quantitative Study on the Exterior Wall Texture of Stone-Built Dwellings in Traditional Villages in China: A Case Study of the Xisuo Village in the Jiarong Tibetan Area. *J. Build. Eng.* **2021**, *42*, 102357. [CrossRef]
22. Liu, C.; Xu, M. Characteristics, and Influencing Factors on the Hollowing of Traditional Villages—Taking 2645 Villages from the Chinese Traditional Village Catalogue (Batch 5) as an Example. *Int. J. Environ. Res. Public. Health* **2021**, *18*, 12759. [CrossRef]
23. Fu, J.; Zhou, J.; Deng, Y. Heritage Values of Ancient Vernacular Residences in Traditional Villages in Western Hunan, China: Spatial Patterns and Influencing Factors. *Build. Environ.* **2021**, *188*, 107473. [CrossRef]
24. Zhao, P.; Miao, Q.; Liu, R.; Song, J. Architectural Style Classification Based on DNN Model. In *Proceedings of the Pattern Recognition and Computer Vision*; Lin, Z., Wang, L., Yang, J., Shi, G., Tan, T., Zheng, N., Chen, X., Zhang, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 505–516.
25. Gao, L.; Wu, Y.; Yang, T.; Zhang, X.; Zeng, Z.; Chan, C.K.D.; Chen, W. Research on Image Classification and Retrieval Using Deep Learning with Attention Mechanism on Diaspora Chinese Architectural Heritage in Jiangmen, China. *Buildings* **2023**, *13*, 275. [CrossRef]
26. Wang, B.; Zhang, S.; Zhang, J.; Cai, Z. Architectural Style Classification Based on CNN and Channel–Spatial Attention. *Signal Image Video Process.* **2023**, *17*, 99–107. [CrossRef]
27. Yang, X.; Song, Z.; King, I.; Xu, Z. A Survey on Deep Semi-Supervised Learning. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 8934–8954. [CrossRef]
28. Van Engelen, J.E.; Hoos, H.H. A Survey on Semi-Supervised Learning. *Mach. Learn.* **2020**, *109*, 373–440. [CrossRef]
29. Yuan, X.; Ou, C.; Wang, Y.; Yang, C.; Gui, W. A Novel Semi-Supervised Pre-Training Strategy for Deep Networks and Its Application for Quality Variable Prediction in Industrial Processes. *Chem. Eng. Sci.* **2020**, *217*, 115509. [CrossRef]
30. Stoller, D.; Ewert, S.; Dixon, S. Adversarial Semi-Supervised Audio Source Separation Applied to Singing Voice Extraction. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*; IEEE: Calgary, AB, Canada, 2018; pp. 2391–2395. [CrossRef]
31. Lisa, F.T.; Hossain, M.Z.; Mou, S.N.; Ivan, S.; Kabir, M.H. Land Cover and Land Use Detection Using Semi-Supervised Learning. In *Proceedings of the 2022 25th International Conference on Computer and Information Technology (ICCIT)*; IEEE: New York, NY, USA, 2022; Volume 10. [CrossRef]
32. Li, X.; Yu, L.; Chen, H.; Fu, C.-W.; Heng, P.-A. Semi-Supervised Skin Lesion Segmentation via Transformation Consistent Self-Ensembling Model. *arXiv* **2018**, arXiv:1808.03887.
33. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C. MixMatch: A Holistic Approach to Semi-Supervised Learning. *arXiv* **2019**, arXiv:1905.02249.
34. Sohn, K.; Berthelot, D.; Li, C.-L.; Zhang, Z.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Zhang, H.; Raffel, C. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *arXiv* **2020**, arXiv:2001.07685.
35. Zhang, B.; Wang, Y.; Hou, W.; Wu, H.; Wang, J.; Okumura, M.; Shinozaki, T. FlexMatch: Boosting Semi-Supervised Learning with Curriculum Pseudo Labeling. *arXiv* **2021**, arXiv:2110.08263.
36. Wang, Y.; Chen, H.; Heng, Q.; Hou, W.; Fan, Y.; Wu, Z.; Wang, J.; Savvides, M.; Shinozaki, T.; Raj, B.; et al. FreeMatch: Self-Adaptive Thresholding for Semi-Supervised Learning. *arXiv* **2022**, arXiv:2205.07246.
37. Zhang, N.; Paluri, M.; Ranzato, M.; Darrell, T.; Bourdev, L. PANDA: Pose Aligned Networks for Deep Attribute Modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Columbus, OH, USA, 2014; pp. 1637–1644.
38. Zou, Q.; Cao, Y.; Li, Q.; Huang, C.; Wang, S. Chronological Classification of Ancient Paintings Using Appearance and Shape Features. *Pattern Recognit. Lett.* **2014**, *49*, 146–154. [CrossRef]
39. Jiang, S.; Shao, M.; Jia, C.; Fu, Y. Consensus Style Centralizing Auto-Encoder for Weak Style Classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Washington, DC, USA, 2016; Volume 30. [CrossRef]
40. Jiang, S.; Shao, M.; Jia, C.; Fu, Y. Learning Consensus Representation for Weak Style Classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2906–2919. [CrossRef] [PubMed]

41. Yu, L.; Liu, J.; Ding, Y.Q. The Spatial Distribution Dataset of 2555 Chinese Traditional Villages. *J. Glob. Chang. Data Discov.* **2018**, *2*, 144–150. [CrossRef]
42. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
43. Tan, M.; Le, Q.V. EfficientNetV2: Smaller Models and Faster Training. *arXiv* **2021**, arXiv:2104.00298.
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Las Vegas, NV, USA, 2016; pp. 770–778.
45. Snedecor, G.W.; Cochran, W.G. *Statistical Methods*, 8th ed.; Iowa State University Press: Ames, IA, USA, 1989.
46. Cover, T.M.; Hart, P.E. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
47. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*; AAAI Press: Portland, OR, USA, 1996; pp. 226–231. Available online: https://cdn.aaai.org/KDD/1996/KDD96-037.pdf (accessed on 14 May 2026).
48. Rodriguez, A.; Laio, A. Clustering by Fast Search and Find of Density Peaks. *Science* **2014**, *344*, 1492–1496. [CrossRef]
49. van der Maaten, L.; Hinton, G. Visualizing Data Using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.