

Article

Reproducible Industrial CT-to-Porosity Metrics with nnU-Net—A Weak Versus Strong Inference Benchmark on Cementitious Slices

Youxi Wang *, Chaowei Sun  and Le Zhang

Shaanxi Key Laboratory of Safety and Durability of Concrete Structures, Xijing University, Xi'an 710123, China; chao_wei_106@126.com (C.S.); 17638730243@163.com (L.Z.)

* Correspondence: 20190074@xijing.edu.cn

Abstract

Porosity-related quantities from industrial X-ray CT depend on segmentation and inference choices. When inference defaults are omitted from the report, void or phase fractions can shift by amounts comparable to slice-to-slice variability. The contribution is metrological rather than architectural: we document a reproducible nnU-Net 2D workflow on Dataset601 CTVoid from semantic labels to slice-wise void fraction, optional two-dimensional connected-component pore summaries, isotropic three-dimensional stacking at 0.058 mm spacing, and spatial axis diagnostics, with region of interest and voxel spacing stated explicitly. The main results pair a weak export policy, defined as a single forward pass per slice without multi-scale fusion or test-time augmentation, with a strong policy that enables multi-scale fusion and flip-based augmentation on the same slice exports and identical weights, on one hundred consecutive slices from one cementitious industrial stack of 1028×1028 pixels. In parallel we report trainer validation on eight named Dataset601 validation cases and mirroring-based test-time augmentation off versus on re-inference on those same cases; case identifiers and the cross-validation split appear in the main text. These quantities answer different questions and must not be substituted for one another or for independent full-stack ground truth. Porosity-related scalars from industrial X-ray CT depend on how segmentation and inference are configured; when defaults are omitted, void fractions can shift by amounts comparable to slice-to-slice variability. For fixed nnU-Net weights on one cementitious industrial slice stack (1028×1028 pixels), we benchmark weak inference (single forward pass, no multi-scale fusion or test-time augmentation) against a strong export policy (multi-scale fusion and flip-based augmentation) on 100 paired slices, and report parallel trainer validation and TTA-off versus TTA-on re-inference on eight Dataset601 hold-out cases. For the industrial dataset, mean void-class IoU between modes is 0.716 (SD 0.043), while strong inference is $\sim 2.6\times$ slower and predicts lower mean void area (2.37% vs. 3.04%). The full weak export gives a 3D void ratio of 2.44% and integrated void volume of 5175 mm^3 . On validation patches, mean void Dice/IoU against the reference are 0.835/0.728, while weak–strong void IoU reaches 0.924 under the nnU-Net-native TTA contrast—quantities that must not be interchanged across domains or definitions. The present benchmark does not include a systematic polymer dosage series, and the study does not equate semantic void with open porosity but provides a reproducible disclosure template relevant to porous and polymer-modified cementitious CT reporting.



Academic Editor: Antonio Caggiano

Received: 18 May 2026

Revised: 8 June 2026

Accepted: 22 June 2026

Published: 25 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: industrial CT; nnU-Net; porous composite; porosity metrics; weak and strong inference; uncertainty; reproducible workflow

1. Introduction

Pore structure influences transport, mechanical coupling, and degradation in cement-based matrices and in hybrid systems where polymeric phases modify the matrix or appear as distinct domains [1,2]. In polymer-modified cementitious materials—latex or redispersible polymer powders, cellulose ethers, superplasticizers, or dispersed fibers—CT-derived void and phase maps are increasingly used to link formulation and processing to microstructure and transport-related response [3–5]. Cross-study or cross-dosage comparisons require a clearly stated measurement chain from reconstruction through segmentation and inference to post-segmentation metrics. Omitting inference settings invites deployment-related bias in reported scalars.

Laboratory and synchrotron CT provide three-dimensional data, but many engineering workflows still operate on two-dimensional slice stacks for storage, visualization, and batch processing. Reviews of X-ray CT for concrete and asphalt summarize acquisition practice and quantitative opportunities [5], and reviews on machine learning in composite micromechanics link imaging-derived descriptors to effective properties [6]. Fully convolutional [7], U-shaped encoder–decoder [8], memory-efficient decoder [9], transformer-augmented [10], and self-configuring training [11] designs now appear routinely in civil-engineering CT pipelines for mesoscale segmentation and post-segmentation quantification [4,12–19]. The materials literature gives less attention to how inference configuration—for example, whether test-time augmentation (TTA) is enabled or multi-scale fusion is applied in a custom export stack—changes the numerical void or phase fractions later cited in durability or transport arguments.

1.1. Related Work

Semantic segmentation matured through pixel-wise fully convolutional learning [7], encoder–decoder designs with strong skip connections [8], and decoder strategies that trade memory for accuracy [9]. Medical imaging extended these ideas with transformer-based hybrids [10], large-scale interactive models and their clinical adaptations [20], and parameter-efficient few-shot tuning of foundation encoders [21]. nnU-Net bundles dataset-aware training, augmentation, and validation into a default workflow for volumetric segmentation [11] and is increasingly used outside biomedical benchmarks.

For cementitious microstructure, early mesoscale work emphasized contrast handling and segmentation quality on 3D X-ray volumes [15]. Later work segments multiple solid phases in concrete CT for mesoscale or micromechanical models [16], applies deep semantic segmentation to strain-hardening cement composites in micro-CT [4], or couples CT voxels with virtual-element reconstructions [22]. Together with X-ray CT studies tied to transport-related deterioration [3], this line of work treats segmentation as the entry point to quantitative morphology, not an isolated labeling task.

Recent concrete and multiphase benchmarks compare FCN, U-Net, DeepLabv3+, and related models on 3D-printed concrete CT with downstream analysis [12]. Lightweight DeepLabv3+ variants with attention target mesostructure and modeling [14]. Dedicated models address permeable concrete mesostructures with reported segmentation performance [19], phase quantification supports the interpretation of ultra-high-performance concrete [18], and fine-grained XCT segmentation targets multiphase building materials [17]. AI-aided interfaces reduce the cost of training masks [13]. Most of this work emphasizes architecture and mIoU-style scores on curated corpora.

Related imaging tasks include surface and volumetric cracks [23–25], asphalt distress [26], sewer defects [27], asphalt mixture structure from X-ray CT [28], and tunnel surveys [29]. Physics differs from bulk void segmentation, but reported performance remains pipeline-dependent (data synthesis, augmentation, post-processing) [25,30].

Taken together, segmentation for construction CT is a crowded front spanning classical CNNs, attention, lightweight design, and foundation-model workflows [7–10,12–14,17–24], while surveys call for transparent inputs and evaluation when learning methods support materials decisions [1,5,6]. Less is said—especially outside the medical-imaging methods literature—about documenting inference-time defaults (TTA, mirroring, multi-scale fusion on exports) and their effect on scalar quantities, such as porosity-like percentages or component counts from fixed weights. Here we foreground paired weak and strong runs, timing, and reference metrics where labels exist [19–24,27–31].

That emphasis responds to recurring omissions that travel together in applications. Post-segmentation metrology from label maps to engineering scalars is often underspecified in reproducible form. Inference-time choices (TTA on/off, multi-scale fusion on slice exports) are seldom reported as part of the uncertainty budget for void-related numbers. Reference agreement (Dice/IoU on held-out labels) is rarely juxtaposed with inter-mode agreement (weak versus strong) on a separate industrial stack, which blurs whether disagreement reflects the learned representation or deployment-specific inference and preprocessing.

1.2. Objective, Scope, and Contributions

This paper does not propose a new backbone network. It isolates a measurement-chain question for nnU-Net-based void analysis. For fixed weights, how do alternative inference policies and reference-evaluation paths change void-class IoU, void fractions, timing, and spatial summaries? The study therefore complements architecture-centric CT segmentation work by foregrounding how semantic void is defined, disclosed, and propagated through post-segmentation metrology rather than by introducing a new network design. Throughout, “void” denotes a discrete supervision target. It is not treated as interchangeable with open porosity or a percolating network without further experiment [2,5,12,14,15]. Qian [32] likewise stresses generalization and dataset shift for concrete image analysis, which supports stating clearly whether metrics come from training patches or from industrial exports.

The paper contributes in three ways. It consolidates a reproducible pipeline from nnU-Net predictions (class 3 in this deployment) to slice-wise and stacked void fractions, optional 2D connected-component summaries, and spatial overlays, sufficient for replication given the same model export and slice geometry. It explains how manual reference segmentations for Dataset601 relate to reporting and why weak–strong IoU on the primary industrial CT stack answers a different question than Dice/IoU against held-out reference labels on Dataset601 validation cases. It reports a controlled benchmark of one hundred paired weak and strong inferences on that stack, with IoU, pixel counts, label fractions, timing, and correlations. The main evidence is metrological (weak versus strong inference and validation paths on one cementitious industrial volume and Dataset601 hold-outs). The work is nonetheless relevant to polymer-in-cementitious CT when several formulations are compared, because it gives a disclosure template and a quantitative sense of how sensitive void-like scalars are to inference defaults.

The same segmentation-to-metrics discipline applies to heterogeneous composites more broadly [1,6], including polymer-rich domains or voids at polymer–matrix interfaces, whenever label semantics are defined consistently.

2. Materials and Data

2.1. Cementitious Specimen and Mix Proportions

The primary industrial CT volume in Section 2.3 is associated with laboratory specimen M1, cast as non-standard 100 mm × 100 mm × 100 mm cubic prisms (100 mm edge length). The geometry was chosen to suit the tomographic field of view and the

present segmentation workflow rather than a particular structural test coupon size. Mix proportioning followed the Chinese industry standard Design Code for Mix Proportion of Hydraulic Concrete (DL/T 5330-2015) [33]: water-to-binder ratio, sand content, and batch quantities of constituents were taken from that procedure. The as-cast mixture proportions are summarized in Table 1 (kg/m³ of concrete), and the materials including cement (Conch Cement Co., Ltd., Wuhu, Anhui, China), fly ash (Henan Borun Foundry Materials Factory, Zhengzhou, Henan, China), sand (Xi'an Tengyuanda Building Materials Co., Ltd., Xi'an, Shaanxi, China), coarse aggregate (Xi'an Tengyuanda Building Materials Co., Ltd., Xi'an, Shaanxi, China), superplasticizer (Jinzhou Hengxiang Chemical Technology Co., Ltd., Jinzhou, Hebei, China). Table 2 lists saturated-surface-dry (SSD) specific gravities and the corresponding absolute volumes per cubic meter of concrete so readers can verify that the batch closes on unit volume.

Table 1. Mix proportions for specimen M1 (kg/m³).

Specimen	Water	Cement	Fly Ash	Sand	Coarse Aggregate	Superplasticizer
M1	200	280	120	820	890	3.2

Table 2. SSD specific gravities and absolute volumes for specimen M1 (per m³ of concrete).

Constituent	Mass (kg/m ³)	SSD γ	Abs. Volume (m ³ /m ³)
Water	200	1.00	0.200
Cement	280	3.15	0.089
Fly ash	120	2.25	0.053
Sand	820	2.62	0.313
Coarse aggregate	890	2.65	0.336
Superplasticizer	3.2	1.1	0.003
Subtotal (solids + water)	2313.2	—	0.994
Air (design balance)	—	—	0.006
Total	—	—	1.000

Absolute volumes were computed as $V_i = m_i / (\gamma_i \times 1000)$. SSD values γ_i were taken from supplier certificates and batch material records for the M1 cast, and the air row balances the sum to a 1.000 m³ per DL/T 5330-2015 proportioning practice.

2.2. Industrial X-Ray CT Acquisition

Volumetric imaging of specimen M1 (Section 2.1) used a MultiscaleVoxel 1000 industrial X-ray computed tomography system (Tianjin Sanying Precision Instruments Co., Ltd., Tianjin, China), shown in Figure 1. The scan was nondestructive and aimed at capturing the internal microstructure of the concrete for subsequent tomographic reconstruction and slice-wise analysis. Stacked quantities in Section 3.3 and the frozen summary in Section 4.2 use 0.058 mm isotropic spacing on the exported analysis grid.

2.3. Primary Industrial CT Stack

The primary industrial CT data comprise 1028 consecutive tomographic slices on a 1028 × 1028 pixel grid, each stored as a grayscale tomographic intensity map together with a co-registered binary void mask obtained from nnU-Net under one fixed weak-inference export setting, including unchanged trained weights and export-script options. Slice-resolved quantities used in this were taken from the archived weak-export statistics bundled with that stack rather than recomputed from raw projections. These archived tables are outputs of the documented export scripts and logs in the reproduction bundle, and they are reported here to avoid redundant recomputation in the manuscript pass, not as

irreproducible ad hoc values. Any third party can regenerate the same summaries from the bundled weak masks and tabulated outputs using the paths listed in the data-availability statement. Table 3 summarizes the distribution of the slice-wise void area fraction ϕ over all 1028 slices; ϕ is defined in Section 3.3.



Figure 1. Multiscale Voxel 1000 industrial CT system.

Table 3. Weak-export pipeline on the primary industrial CT stack for 1028 slices.

Metric	Value (%)
Mean ϕ	2.44126
Std. dev.	0.43041
Min/Max	1.53645/3.58787
Median	2.38152
p10/p25/p75/p90	1.91638/2.16284/2.74775/3.08159

The weak–strong benchmark uses the first one hundred slices that are present in both weak and strong runs. The full stack (1028 slices) supports weak-only distributional and 3D summaries. The hundred-slice subset is a separate, fully paired weak–strong cohort, and the first 100 stems are present in both export runs after lexicographic filename sort (the same deterministic rule implemented in the bundled comparison script), chosen for transparent replication rather than random subsampling. Intermediate factorial inference settings (e.g., multi-scale only or TTA only on the industrial export path) are outside the present scope. Per-slice and aggregate metrics were written to the benchmark logs summarized in Tables 4 and 5.

The Pearson correlations r (weak void px, strong void px) and r (weak void %, strong void %) are numerically identical because every slice shares the same 1028×1028 grid, so void area percent is a fixed positive rescaling of void pixel count.

Per-slice wall-clock times in Tables 4 and 5 were recorded during GPU-accelerated inference on the primary-stack export path and transcribed from the archived benchmark log without post hoc adjustment. Table 4 reports cohort-level timing statistics (mean per slice, strong-to-weak ratio, and totals over 100 slices), and Table 5 reports label-mass and correlation quantities of the same paired cohort. All timed runs used the same hardware–software pairing and identical trained weights, while weak and strong modes differ only in the export-script inference settings defined in Section 3.1. Timing was not standardized for

the Dataset601 re-inference workflow. The nnU-Net was used in accordance with version 2 library conventions, and the deployed weights match the bundled Dataset601_CTVoid training configuration. In the primary-stack integer label maps used here, void corresponds to class index 3 under the semantic mapping implemented in the custom export.

Table 4. Weak versus strong inference for first 100 paired slices on the primary industrial CT stack.

Quantity	Value
n slices	100
Mean void IoU	0.71637
Std. dev. IoU	0.04330
IoU min/max	0.61775/0.78458
IoU p10/p50/p90	0.65073/0.72378/0.76834
Mean weak time/slice (s)	33.808
Mean strong time/slice (s)	87.689
Mean ratio strong/weak	2.594
Total weak time (s)	3380.761
Total strong time (s)	8768.862

Table 5. Label-mass summary of the same 100 slices.

Quantity	Value
Mean weak void pixels	32,077.6
Mean strong void pixels	25,035.6
Mean Δ void px (strong – weak)	–7042.1
Mean exclusive weak-only void px	8305.7
Mean exclusive strong-only void px	1263.7
Mean weak void % (class 3)	3.0354
Mean strong void % (class 3)	2.3690
Mean Δ void % (strong – weak)	–0.6664
r (weak void px, strong void px)	0.80614
r (weak void %, strong void %)	0.80614
r (gray on weak void, gray on strong void)	–0.02534
r (exclusive weak-only px, exclusive strong-only px)	0.58917

2.4. Reference Segmentations and Training Data for Dataset601_CTVoid

The study bundle contains reference segmentations and imaging patches used to train nnU-Net, namely manual label patches, paired 16-bit image patches, and documentation that records the label convention (void, paste, coarse aggregate semantics in that export). The integer label map after conversion into the Dataset601 nnU-Net dataset format must be stated in the final text, because preprocessing may remap indices relative to raw patch files. Within this study bundle, void is indexed as class 3 in all integer export maps analyzed here; transferring the checkpoint to another project requires checking the dataset meta-data and export mapping because indices are deployment-specific rather than universal across datasets.

These data supply supervision for learning only. Held-out validation metrics for Dataset601 come from the nnU-Net trainer on $n = 8$ validation cases—CTVOID_0002, CTVOID_0003, CTVOID_0005, CTVOID_0019, CTVOID_0021, CTVOID_0025, CTVOID_0028, and CTVOID_0032—trained under the first of the library’s five cross-validation splits. Each case is scored against preprocessed reference segmentations on the validation split, not against every slice of the primary industrial CT stack. These eight identifiers are the fold-0 validation hold-outs assigned by nnU-Net, not an author-selected sample. Metrics on this small fold characterize reference agreement on the training/validation distribution

only and must not be read as unbiased generalization to the industrial volume or to other cross-validation splits. Table 6 summarizes the aggregate statistics.

Table 6. Dataset601_CTVoid trainer validation: mean Dice and IoU by class for eight held-out cases.

Metric	Paste	Coarse	Void
Mean Dice	0.940578	0.954377	0.835401
Mean IoU	0.889969	0.913870	0.728163

3. Methods

3.1. Segmentation and Inference Modes

Let I denote a grayscale slice. Weak and strong inference each produce a discrete semantic label map on the same $H \times W$ grid, and both use the same trained weights.

$$Y^{(w)}, Y^{(s)} \in \{0, 1, 2, 3\}^{H \times W}. \quad (1)$$

Binary void masks for class index 3 are

$$M^{(w)} = 1[Y^{(w)} = 3], \quad M^{(s)} = 1[Y^{(s)} = 3]. \quad (2)$$

Weak inference applies a single forward pass per slice without multi-scale fusion or TTA. Strong inference enables multi-scale factors and flip-based TTA. On the primary-stack export path, weak and strong bracket a deliberate minimal-versus-maximal deployment contrast on identical trained weights, while weak disables both multi-scale fusion and TTA, and strong applies multi-scale fusion (factors 1.0, 0.75, 0.5 with fixed weights 0.15, 0.3, 0.55 in the bundled script) and flip-based TTA jointly. Their individual contributions are not factorially separated in the industrial stack, and intermediate combinations are reserved for future ablation. Section 3.6 defines a second, nnU-Net-native weak/strong pair used only on Dataset601 validation patches. Qualitative and quantitative outcomes for the primary-stack export pipeline are reported in Section 4.1 and Tables 4 and 5 in Section 2.3.

3.2. Pairwise Metrics and Timing

For each paired slice, we compute the intersection over union (IoU) between the void masks $M^{(w)}$ and $M^{(s)}$ from Section 3.1, together with void pixel counts, exclusive weak-only and strong-only void pixels, per-class label percentages, mean gray values on void pixels, and wall-clock times.

$$\text{IoU}(M^{(w)}, M^{(s)}) = \frac{|M^{(w)} \cap M^{(s)}|}{|M^{(w)} \cup M^{(s)}|}. \quad (3)$$

Per-slice records for the hundred-slice benchmark are read from the frozen weak–strong comparison table.

3.3. Slice-Wise and Stacked Void Fraction

Let N be the number of void-class pixels on a slice, with image height H and width W . The slice-wise void area fraction ϕ (percent) is

$$\phi = 100 \times \frac{N}{HW}. \quad (4)$$

Stacked 3D void ratio and integrated void volume use the weak-export void mask on the same isotropic 0.058 mm voxel grid as the slice exports. These 3D quantities are

post-segmentation summaries of stacked two-dimensional nnU-Net slice exports under the deployed 2D trainer configuration, not outputs of a separate native three-dimensional nnU-Net inference on the full volume. The industrial workflow batch-processes 1028 grayscale slice images. To enable metrological comparison between weak and strong inference, a fixed 2D checkpoint and exported geometry were retained. Tabulated stack summaries and slice-order plots appear in Section 4.2.

3.4. Two-Dimensional Equivalent Diameter Distribution

On each 8-connected component of a void mask, let A be the pixel area of the component, px_mm the in-plane millimeters per pixel, and $d_{eq,mm}$ the equivalent diameter (mm). Then

$$d_{eq,mm} = 2\sqrt{\frac{A}{\pi}} px_mm. \quad (5)$$

Components are classified into configurable small/medium/large bins using the 2D pore-size analysis. Supplementary pore-bin tables should be generated only from masks exported for the stated geometry and voxel spacing.

3.5. Spatial Diagnostics

To examine through-stack structure and lateral edge effects, void masks from the weak export were trimmed by eight pixels on each x -side before averaging. Slice-wise void fraction was retained as a function of sort index normal to z , and marginal profiles were obtained by averaging void fraction over x and over y on the trimmed domain. Tabulated profiles and the spatial summary file used for numerical ranges are archived with the reproduction bundle. Section 4.2 visualize the same diagnostics and the marginal distribution of ϕ over all weak-export slices.

3.6. Hold-Out Validation: TTA off Versus on for Eight Cases

Let P and R denote binary prediction and reference masks on the same pixel grid (here, void as class index 3). Dice and IoU are

$$\text{Dice}(P, R) = \frac{2|P \cap R|}{|P| + |R|}, \quad \text{IoU}(P, R) = \frac{|P \cap R|}{|P \cup R|}, \quad (6)$$

where $|\cdot|$ counts foreground pixels.

For each of CTVOID_0002, CTVOID_0003, CTVOID_0005, CTVOID_0019, CTVOID_0021, CTVOID_0025, CTVOID_0028, and CTVOID_0032, nnU-Net v2 prediction was re-run using the two-dimensional configuration and the same model checkpoint as in Table 6 (first cross-validation split of the 2D trainer) twice, first with mirroring-based test-time augmentation (TTA) disabled and second with TTA enabled at nnU-Net library defaults. Inputs were the preprocessed case volumes distributed with Dataset601 for that identifier. Outputs were reduced to a single 2D evaluation plane whenever a singleton depth axis remained after collapsing leading singleton dimensions, following one fixed rule applied identically to every case. For void (class index 3) we computed per-case and mean Dice and IoU against the stored reference segmentation supplied for training, together with void IoU between the weak and strong prediction maps. This nnU-Net-native weak/strong pair is not interchangeable with the primary-stack export policies in Section 3.1, where multi-scale fusion on slice exports further changes the prediction map.

3.7. Use of Generative AI in Manuscript Preparation

During the preparation of this manuscript, generative AI tools were used in a limited and auxiliary manner. Specifically, they were applied to language polishing, sentence-

structure optimization, and reference formatting in selected non-core paragraphs, and as an auxiliary aid in describing the code- and algorithm-related parts of the preprocessing pipeline. It should be emphasized that AI tools were not used to substantially generate or replace the core research content (including the research framework and code implementation) of the entire manuscript. The authors reviewed, edited, and take full responsibility for the final text.

4. Results

This section reports numerical outcomes for the materials and primary-stack data in Section 2 under the procedures in Section 3. Section 4.1 presents paired weak and strong behavior of one hundred industrial slices. Section 4.2 summarizes the full weak-export stack, distribution of slice-wise void fraction, isotropic three-dimensional void ratio, and axis-resolved profiles after eight-pixel lateral trimming. Section 4.3 reports Dataset601 trainer validation and paired re-inference with test-time augmentation disabled versus enabled at nnU-Net library defaults on eight validation cases. Section 4.4 synthesizes the three empirical strands and stresses that their metrics answer different questions and must not be conflated. Section 5 gives interpretation, limitations, and reporting recommendations.

4.1. Paired Weak and Strong Inference on One Hundred Primary-Stack Slices

The weak–strong benchmark cohort has a fixed size of

$$n = 100. \quad (7)$$

The weak–strong benchmark comprises the first one hundred paired slices from the primary industrial stack for which both export-based predictions exist, selected by lexicographic filename order as stated in Section 2.3. Per-slice definitions and metrics follow Sections 3.1 and 3.2, and cohort-level timing and label-mass summaries are given in Tables 4 and 5. Inferential summaries here describe slice-to-slice variability within specimen M1, not population-level significance across multiple specimens or formulations.

Figure 2 shows one representative in-plane field with co-registered CT intensity and weak versus strong semantic maps on the same grid, using fixed windowing and color scales that match the conventions in Figures 3 and 4.

Mean void-class IoU between the two void masks is 0.716, with standard deviation 0.043, extrema 0.618 and 0.785, and an inter-decile band of approximately 0.651–0.768 (Table 3). In aggregate, the two masks overlap substantially, yet slice-resolved disagreement is non-negligible and varies across the cohort. Figure 3 reads by rows. Panels a–c summarize cohort-level agreement and cost: the marginal distribution of void IoU (a), weak versus strong wall time per slice with the identity line (b), and the histogram of the strong-to-weak time ratio (c). Panels d–f link disagreement to void mass and to area fractions attributed to the void class: weak versus strong void pixel counts with points colored by IoU (d), the histogram of the signed pixel surplus for strong minus weak (e), and weak versus strong void area percent with the identity line (f). Panels g–i emphasize systematic shift and slice ordering: a Bland–Altman-style plot of void percent with mean difference and approximate 95% limits (g), void IoU along the sorted slice index used in the benchmark (h), and exclusive weak-only versus strong-only void pixel counts (i). In scatter panels b, f, and i, annotated r is the Pearson correlation coefficient across $n = 100$ slices, while (b) is weak versus strong wall time (s); (f) is weak versus strong void area % (class 3); and (i) is exclusive weak-only versus exclusive strong-only void pixels.

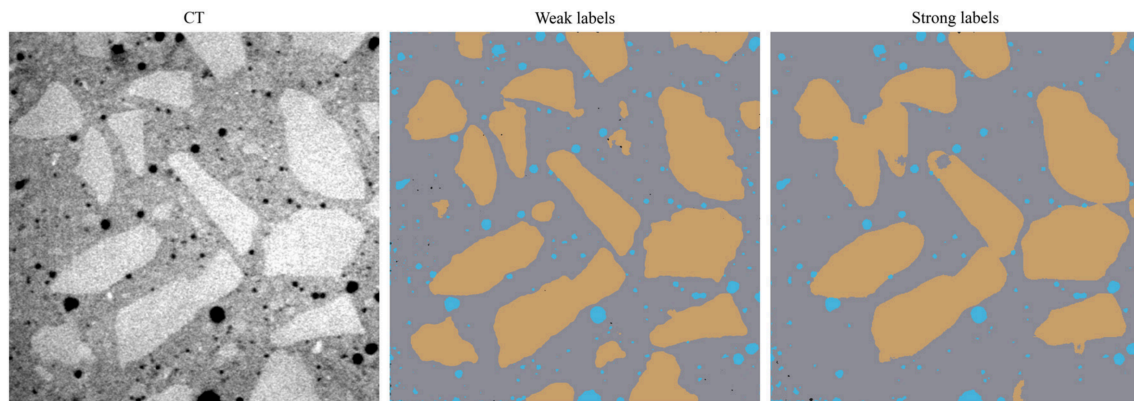


Figure 2. Representative CT slice with weak versus strong segmentation (the color orange = coarse, blue = void, gray = paste).

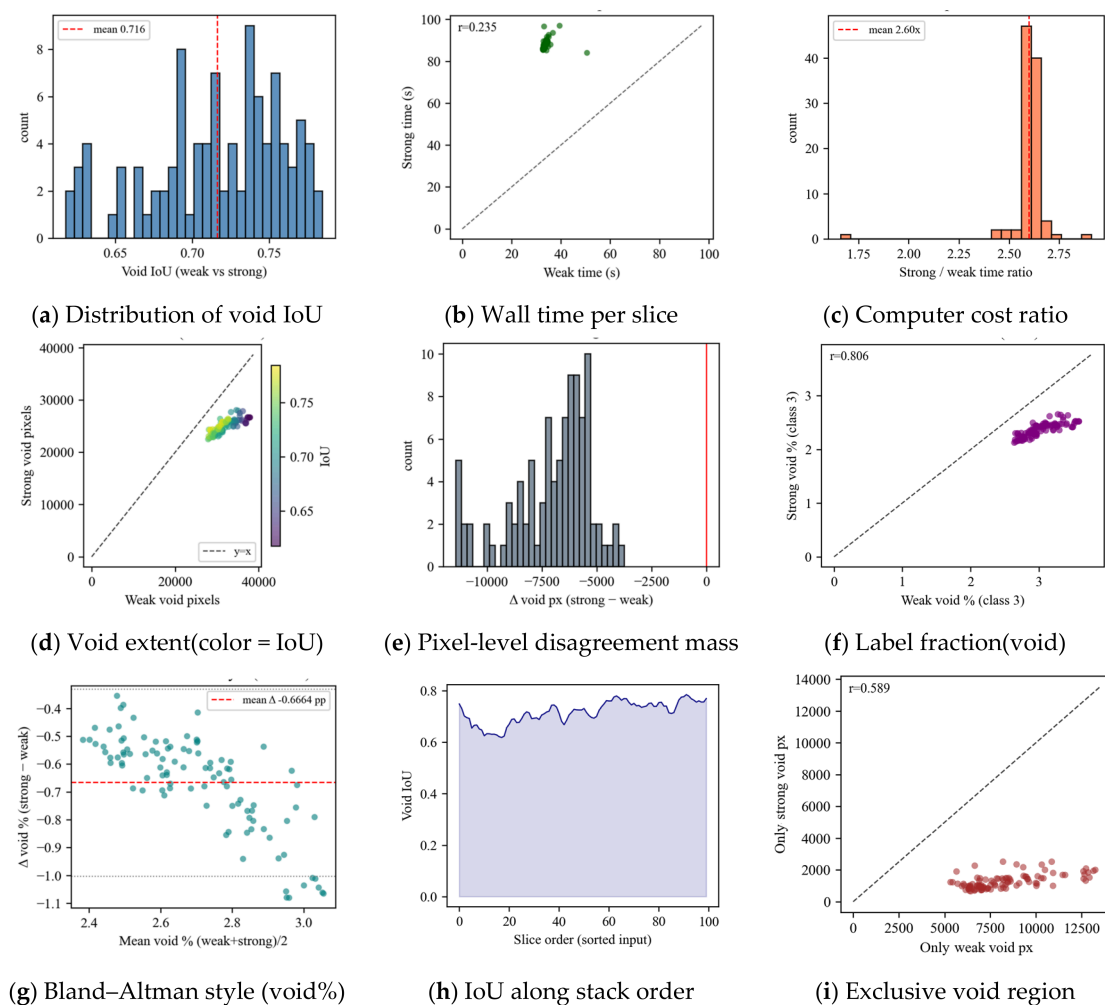


Figure 3. Weak versus strong inference (multi-panel overview).

Figure 4 complements Figure 3 on the same cohort. Panels a and b scatter weak versus strong area fractions for paste and coarse, respectively, with the identity line and correlation annotated in each panel. Panel c gives the cohort distribution of the slice-wise change in void area fraction (strong minus weak). Panel d compares mean CT intensity restricted to void pixels under weak versus strong. Together with Table 5, it motivates the near-zero correlation between gray-on-void summaries across modes reported in the next paragraph.

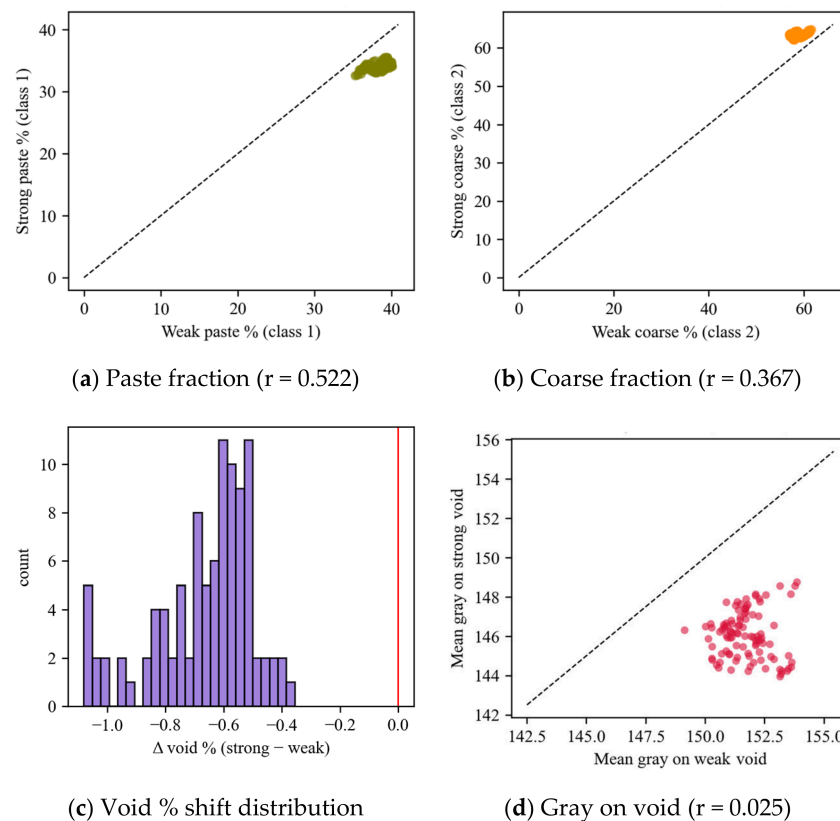


Figure 4. Weak versus strong inference (class balance and intensity on void).

Mean wall-clock time per slice is 33.8 s (weak) versus 87.7 s (strong), so strong inference is $2.59\times$ slower on average for this hardware–software pairing (Table 4). Totals over 100 slices exceed 0.9 h (weak) versus 2.4 h (strong), so enabling multi-scale fusion and augmentation carries a repeatable computational cost as well as a statistical one.

Strong mode assigns fewer void pixels on average than weak (25.0×10^3 versus 32.1×10^3 mean pixels per slice), corresponding to mean void area fractions of 2.37% versus 3.04% for class 3—about 0.67 percentage points lower in the mean for strong (Table 5). Exclusive weak-only void pixels exceed strong-only void pixels on average, consistent with strong removing more void voxels than it adds relative to weak. Pearson correlations between weak and strong void masses are approximately

$$r_{px} \approx 0.81, \quad r_{\%} \approx 0.81, \quad (8)$$

so slice ranking by void extent is largely preserved even when absolute void mass shifts. The equality $r_{px} = r_{\%}$ follows from the common 1028×1028 slice grid (Table 5).

The Pearson correlation of mean gray evaluated on void pixels (weak versus strong) is approximately

$$r_{gray} \approx -0.03. \quad (9)$$

Mean gray on void pixels is therefore essentially uncorrelated between weak and strong in this cohort (Table 5, Figure 4), and simple intensity summaries on the segmented void mask do not track weak–strong disagreement in this dataset.

4.2. Full-Stack Slice Statistics, Isotropic 3D Stacking, and Spatial Shape

Slice-wise void fraction ϕ and isotropic three-dimensional stacking are specified in Section 3.3, and lateral boundary trimming and marginal void-fraction profiles follow Section 3.5. Quantitative entries and the figures in this subsection (Table 3 for slice-wise ϕ

over 1028 slices, Table 7 for the isotropic 3D stack summary; Figures 5–7) reproduce tabulated outputs from the weak-only segmentation export archived with this study, without repeating inference on the primary-stack slices.

Table 7. Stacked weak void volume for the primary stack.

Quantity	Value
Slices in stack	1028
In-plane/slice spacing (mm)	0.058 (isotropic in file)
3D void ratio (%)	2.441261
Total void volume (mm ³)	5174.610

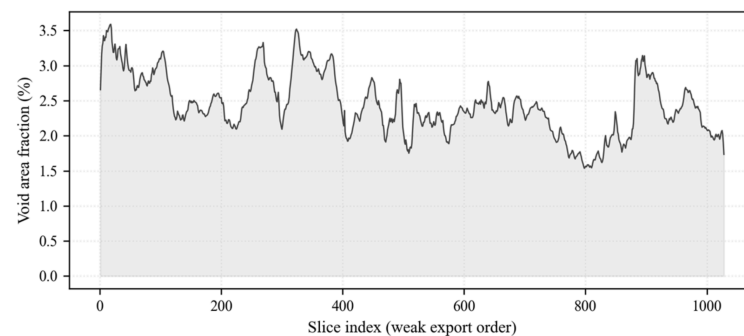


Figure 5. Slice-wise void area fraction versus slice index.

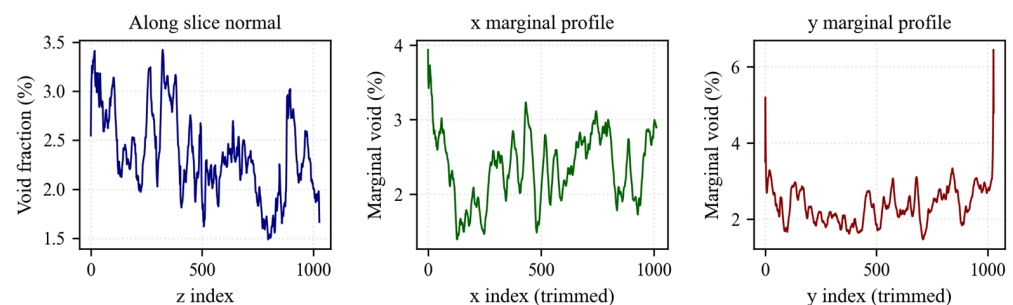


Figure 6. Spatial void profiles after x-side trimming.

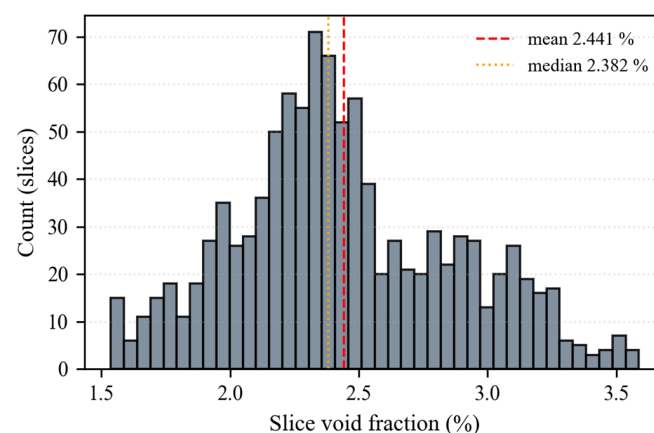


Figure 7. Histogram of slice-wise void fraction (primary stack).

Across all 1028 weak-export slices (Table 3, Figures 5 and 7), mean slice-wise void fraction is 2.44% with a standard deviation of 0.43% and an empirical range from about 1.54% to 3.59%. Quartiles and the upper decile indicate substantial slice-to-slice variability relative to the mean. The 90th percentile of ϕ satisfies approximately

$$p90(\phi) \approx 3.08\%. \quad (10)$$

Figure 5 shows how void fraction varies along the sort order. Figure 7 gives a unimodal, right-skewed marginal distribution of ϕ over all 1028 weak-export slices on specimen M1 (primary industrial CT stack), consistent with the tabulated percentiles in Table 3.

Stacking the weak void masks at 0.058 mm isotropic spacing (Table 7) yields a 3D void ratio of 2.441% and an integrated void volume of 5175 mm³ for the full stack. After 8-pixel trimming on each x -side for axis diagnostics (Figure 6), the global void ratio from the spatial summary drops to 2.361%, and slice-resolved profiles along z span roughly 1.49–3.42% of the void fraction. Through-stack variation is therefore comparable in magnitude to the mean-level shift between weak and strong on the hundred-slice benchmark. Marginal profiles in x and y (Figure 6) indicate an anisotropic void fraction structure after trimming, including edge-adjacent bands at the reported y marginal maximum (6.45% in the archived spatial summary produced under Section 3.5).

The industrial stack evidence above is entirely export-based on specimen M1. The following subsection returns to the bundled Dataset601 supervision, Table 6 records trainer validation against the distributed reference.

4.3. Trainer Validation on Dataset601 and Paired Re-Inference on the Same Eight Cases

Table 6 (Section 2.4) aggregates nnU-Net’s validation predictions against preprocessed reference segmentations for eight held-out cases. Mean Dice/IoU for void are 0.835/0.728. The foreground mean (nnU-Net aggregate) is Dice 0.910118 and IoU 0.844001. For void across the eight cases, Dice ranges from 0.709880 to 0.972753 (std 0.098894) and IoU from 0.550244 to 0.946951 (std 0.146082). Agreement with reference on held-out patches is moderate to good for void but varies by case. Mean void IoU (0.728) is numerically close to mean weak–strong void IoU (0.716, Table 4), yet the two quantities are not interchangeable: one scores prediction against ground truth on Dataset601 validation patches, whereas Table 4 compares two export-based inference modes on primary-stack slices.

Table 8 uses the nnU-Net-native weak/strong pair from Section 3.6 (TTA off versus default TTA on). Mean Dice is 0.832 (weak) versus 0.835 (strong), and mean IoU is 0.723 versus 0.728 for void, so strong is slightly closer to the stored reference on average. The gains are small and not uniform (for CTVOID_0005, strong Dice falls below weak). Weak and strong prediction maps on these patches still agree at mean void IoU of 0.924—much higher than the 0.716 mean weak–strong void IoU on the primary-stack export benchmark (Table 4) because patch geometry, class semantics, and the operational meaning of “weak/strong” differ from Section 3.1, and Dataset601 patches follow the training distribution rather than full industrial slices.

Table 8. Eight Dataset601 validation cases: void Dice and IoU against the reference for test-time augmentation off (weak) versus library-default on (strong).

Case	Dice Weak	IoU Weak	Dice Strong	IoU Strong	IoU Weak–Strong (Void)
CTVOID_0002	0.724782	0.568359	0.732075	0.577381	0.927894
CTVOID_0003	0.711377	0.552045	0.730627	0.575581	0.916364
CTVOID_0005	0.726601	0.570600	0.709068	0.549268	0.871589
CTVOID_0019	0.893855	0.808081	0.906788	0.829472	0.922914
CTVOID_0021	0.972168	0.945844	0.972753	0.946951	0.965705
CTVOID_0025	0.904196	0.825145	0.911811	0.837916	0.917271
CTVOID_0028	0.864481	0.761309	0.858089	0.751451	0.919568
CTVOID_0032	0.857320	0.750271	0.859988	0.754367	0.952179
Mean	0.831848	0.722707	0.835150	0.727798	0.924185

4.4. Integrated Synthesis

The Section 4 results are structured along three complementary lines of evidence, supported by Tables 3–8 and Figures 2–7. Table 1 (Section 2.1) documents batch mixture proportions for specimen M1 as auxiliary materials information for the scanned coupon and lies outside the segmentation evaluation. The first line combines the representative in-plane triptych (Figure 2) with cohort-level weak–strong metrics on industrial slice exports (Tables 4 and 5) and the associated multi-panel summaries (Figures 3 and 4). The second line reports through-stack slice-wise void fraction, isotropic three-dimensional void summaries, and spatial profiles after lateral trimming (Tables 3 and 7; Figures 5–7). The third line gives reference-scored validation on the eight Dataset601 hold-out cases (Tables 6 and 8). Mean void IoU against the stored reference (Table 6, 0.728) and mean void IoU between weak and strong predictions on the hundred-slice primary-stack benchmark (Table 4, 0.716) are numerically close, yet they quantify distinct experimental contrasts—agreement with bundled labels on preprocessed validation patches versus overlap between export-based prediction modes on industrial slices under different weak/strong definitions—and must not be treated as interchangeable. Section 5 discusses reporting practice and cautions against conflating semantic void with physical porosity or with deployment-specific inference defaults that are left unstated in the main text.

5. Discussion

5.1. Synthesis

The three comparisons differ in supervision geometry, in the operational definitions of weak versus strong inference, and in imaging domain; treating them as interchangeable would conflate distinct experimental questions into a single summary scalar. As emphasized in Section 4.4, numerically similar void IoU values can correspond to different contrasts. The paragraphs that follow develop implications for transparent reporting and for reading export-based weak–strong benchmarks alongside Dataset601 validation metrics.

For porous matrices, polymer-modified binders, and multiphase composites, void- or phase-related scalars from CT depend jointly on the model checkpoint, inference configuration (including mirroring-based test-time augmentation and multi-scale fusion), the class index used for void in each workflow, whether metrics are computed on two-dimensional slice exports or native three-dimensional volumes, and the presence, geometry, and pre-processing of any reference segmentations. Where polymer-rich regions or interfacial voids overlap in gray level with hydrates or aggregates, that dependence strengthens, and identical inference settings become important for fair comparison across formulations or curing ages. Global thresholding and pixel counting alone are an insufficient description of the chain when learning-based segmenters are used. This study quantifies how sensitive void-class scalar summaries are to the preceding choices when they are left unstated in the main text or repository.

This positioning is consistent with CT reviews for concrete and composite micromechanics, which stress transparent acquisition-to-quantification chains rather than headline segmentation scores alone [5,6], and with self-configuring segmentation workflows such as nnU-Net, where trainer validation on held-out cases is informative but does not by itself specify deployment behavior under full industrial exports [11,32]. Relative to cementitious deep-learning CT studies that foreground architecture and mIoU on curated corpora [4,12–19], the present work keeps the checkpoint fixed and varies inference disclosure paths, which is closer to a metrological than an architectural comparison.

5.2. Weak Versus Strong Is Not a Single Universal Curve

The primary-stack export pipeline defines strong inference through multi-scale fusion and flip-based augmentation in the bundled export tooling, whereas weak inference uses a single forward configuration. On one hundred consecutive slices, mean void IoU between the two masks is about 0.72 (Table 4), yet strong predicts fewer void pixels on average than weak (Table 5). That pattern fits a sharper decision boundary under the heavier policy. Void voxels are removed in some regions and only partly recovered elsewhere, so global void fraction drops while overlap remains substantial.

On the primary stack, multi-scale fusion and TTA are enabled together in the bundled export script, whereas Dataset601 re-inference (Table 8) isolates TTA off versus on under nnU-Net defaults without that export-script multi-scale fusion.

The Dataset601 validation re-runs instead define weak and strong through TTA disabled versus default mirroring-based TTA. There, strong is marginally closer to the stored training segmentations for void on average, and weak–strong void overlap is much higher (mean IoU about 0.92) than on the primary stack. The gap relative to Table 4 is therefore expected because definitions of weak/strong differ, the image domains differ (training patches versus full industrial slices), and effective class geometry differs, so the trade-off among accuracy, consistency, and cost moves. The case of CTVOID_0005 shows non-monotonicity at the case level (strong Dice below weak), so endorsing the maximal inference policy without case-specific validation is not supported by these data.

Reported cementitious CT segmentation studies often report strong overlap with manual or curated references on training-like patches or mesoscale volumes [4,14–17] yet seldom tabulate how alternative inference defaults on production slice exports change void fractions and runtime on the same weights. The industrial-stack mean weak–strong void IoU of 0.72 together with a 0.67 percentage-point mean shift in void area under the heavier export policy illustrates that deployment choices can move porosity-like scalars by magnitudes comparable to slice-to-slice variability (Table 3), even when masks remain substantially correlated. The much higher patch-level weak–strong IoU (0.92, Table 8) is therefore not contradictory, and it aligns with expectations under dataset shift and changed operational definitions of weak and strong, as emphasized in concrete image-analysis generalization work [32].

5.3. Reference Labels, Bundled Validation, and What Remains Open

Manual annotations in the Dataset601 training bundle define supervision semantics. They are required for training but do not transfer to every downstream folder without registration, cropping, and resolution alignment. Table 6 reproduces the trainer’s validation summary for the eight Dataset601 cases in Section 2.4 against preprocessed reference segmentations. Table 8 scores stored training segmentations after paired re-inference. Together they characterize only the training distribution. They address whether, for those eight validation cases, the network matches the stored reference closely and whether native TTA helps slightly on average.

These validation statistics do not identify which primary-stack inference mode is closer to an independent human trace on each industrial slice, nor do they establish open porosity or percolation. They report only agreement with a semantic label map under Dataset601’s conventions. On the industrial hundred-slice cohort, mean gray evaluated on void pixels under weak versus strong is nearly uncorrelated across slices (Table 5, Figure 4), so simple intensity summaries on the segmented void mask do not track inter-mode mask disagreement there. Claims about transport, mechanics, or durability still call for orthogonal validation (e.g., repeat scans, mercury intrusion where appropriate, or stereology) alongside segmentation metrics.

Mean void Dice/IoU against the bundled reference on the eight validation cases (Table 6; void IoU 0.728) are broadly comparable in order of magnitude to void-segmentation performance reported for multiphase cementitious CT and related mesostructure pipelines [4,12,14,17,19], but they characterize agreement on preprocessed training-domain patches rather than on every slice of the industrial volume. Juxtaposing Table 6 with Table 8 makes it explicit that native TTA yields only marginal average gains against the stored reference while preserving very high inter-mode overlap on patches—behaviors that should not be extrapolated to the export-based weak/strong pair in full industrial slices (Tables 4 and 5). This separation parallels calls in composite and concrete CT surveys for stating which geometry, label convention, and inference path each reported scalar refers to [5,6].

5.4. Reporting Recommendations and Limitations

CT-based composite papers should report at a minimum which model checkpoint and which cross-validation split (or explicit list of held-out case IDs) were used, whether TTA and multi-scale fusion were used (with non-default weights if any), the label index used for “void” in figures, whether metrics come from 2D slice exports or native 3D volumes, and whether held-out reference metrics exist, on which geometry they were computed, and how they relate to the production inference path. Segmentation, post-processing, and inference defaults are part of the method, not unspecified preprocessing, in line with transparency expectations for learning-based CT pipelines.

Such disclosure is especially relevant for polymer-modified and multiphase cementitious systems, where CT-derived void or phase maps increasingly support transport and durability interpretation [1–3] and where cross-formulation comparison requires identical inference and post-segmentation settings rather than architecture-only benchmarking [6,12,14]. The checklist above extends review-level recommendations for concrete CT and composite micromechanics [5,6] to nnU-Net-style deployments in which weak versus strong choices are easy to omit from the published record.

The industrial benchmark uses one contiguous block of slices from a single volume. That design supports a controlled comparison of inference policies but is not a multi-specimen or multi-site study, and external validity beyond this acquisition is not established. The basic specimen geometry (non-standard 100 mm cube, Table 1 mix) is stated in Section 2.1; Section 2.2 identifies the industrial CT platform (Figure 1). The detailed source, detector, and reconstruction settings are not tabulated here; downstream metrics use the exported analysis grid at 0.058 mm spacing. Mix proportions for M1 are given in Table 1. Dataset601 validation here uses only the eight cases (CTVOID_0002, CTVOID_0003, CTVOID_0005, CTVOID_0019, CTVOID_0021, CTVOID_0025, CTVOID_0028, CTVOID_0032) under the re-inference procedure in Section 3.6 and the first of five nnU-Net cross-validation splits, so case-level variability and behavior across other splits are only partly characterized. Wall-clock time was logged systematically for the primary-stack benchmark, whereas the validation re-inference workflow prioritized metric reproducibility over standardized timing across environments. Integer class indices in nnU-Net exports must stay aligned with the Dataset601 semantic keys (void, paste, coarse) and with the void-class index in downstream processing.

6. Conclusions

This paper documents a reproducible, script-linked route from fixed nnU-Net weights to void-related slice, stack, and axis-resolved summaries on one industrial cementitious CT export, together with the bundled Dataset601 supervision bundle. Evidence is organized in three non-substitutable strands: export-based weak versus strong behavior concerning

one hundred primary-stack slices, through-stack distribution and isotropic 3D stacking on the full weak export, and reference-scored validation plus native TTA-off versus TTA-on re-inference on eight named Dataset601 validation cases. Concerning the industrial benchmark, mean weak–strong void IoU is 0.716 with a $2.59\times$ mean runtime ratio and lower mean void area under the strong export policy than under weak; for the protocol, strong is only marginally closer to stored training segmentations than weak on average, while weak and strong predictions on those patches still overlap strongly (mean void IoU 0.924)—showing that numerically similar IoUs can attach to different experimental meanings.

The broader lesson is metrological. Semantic void from a trained segmenter is a useful but deployment-sensitive quantity; it should not be equated with open porosity, silently transferred across checkpoints or export pipelines, or conflated with reference metrics computed on another domain or under another definition of weak and strong. For polymer-modified or multiphase cementitious CT where several formulations or curing ages are compared, the same disclosure checklist supports fair cross-paper comparison when void-like fractions enter transport or durability reasoning.

Future work could extend the reporting structure to polymer-matrix or interfacial void semantics once labels and acquisitions are aligned across specimens, add co-registered human or physical references on the industrial volume where feasible, repeat the TTA comparison on additional cross-validation folds, and perform factorial ablation of multi-scale fusion and TTA on the industrial export path.

Author Contributions: Methodology, Y.W.; Software, Y.W.; Investigation, L.Z.; Resources, C.S.; Writing—original draft, Y.W.; Funding acquisition, C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shaanxi Provincial Department of Education’s Special Plan for Local Services Project (22JE019), the Natural Science Foundation Research Project of Shaanxi Province (2025JC-YBMS-511), the Special Fund for High-level Talents of Xijing University (XJ24B12), the Service-Oriented Local Development Special Program Project of the Education Department of Shaanxi Province (25JE019).

Data Availability Statement: All data included in this study are available upon request by contact with the corresponding author.

Acknowledgments: Generative AI tools were used in a limited manner during manuscript preparation for language polishing, sentence-structure optimization, and reference formatting in non-core paragraphs, and as an auxiliary aid in describing code- and algorithm-related sections. AI tools were not used to substantially generate or replace the core research content (including the research framework and code implementation). All authors reviewed and take full responsibility for the final text.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Y.; Makklang, S.; Li, L.; Chen, Z. Integrating domain knowledge in AI-driven geopolymers concrete property prediction system: A comprehensive survey. *Eng. Appl. Artif. Intell.* **2026**, *176*, 114761. [\[CrossRef\]](#)
2. Liu, F.; Tang, R.; Ma, W.; Yuan, X. Analysis on frost resistance and pore structure of phase change concrete modified by Nano-SiO₂ under freeze–thaw cycles. *Measurement* **2024**, *230*, 114524. [\[CrossRef\]](#)
3. Ren, D.; Wang, Z.; Xu, J.; Luo, W.; Su, S.; Li, F.; Ai, C.; Kong, L. Microscopic mechanisms and chloride resistance performance of silane coupling agent self-assembly on reinforced concrete utilizing X-ray CT technology: A comprehensive investigation. *J. Build. Eng.* **2023**, *80*, 107992. [\[CrossRef\]](#)
4. Lorenzoni, R.; Curosu, I.; Paciornik, S.; Mechtcherine, V.; Oppermann, M.; Silva, F. Semantic segmentation of the micro-structure of strain-hardening cement-based composites (SHCC) by applying deep learning on micro-computed tomography scans. *Cem. Concr. Compos.* **2020**, *116*, 103551. [\[CrossRef\]](#)

5. du Plessis, A.; Boshoff, W.P. A review of X-ray computed tomography of concrete and asphalt construction materials. *Constr. Build. Mater.* **2019**, *199*, 637–651. [\[CrossRef\]](#)
6. Huang, K.; Wang, B.; Guo, L. Application of machine learning in the micromechanics of composites: A review. *Compos. Part A Appl. Sci. Manuf.* **2026**, *191*, 109808. [\[CrossRef\]](#)
7. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [\[CrossRef\]](#)
8. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, 18th International Conference, Munich, Germany, 5–9 October 2015, Proceedings, Part III*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [\[CrossRef\]](#)
9. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Chen, J.; Mei, J.; Li, X.; Lu, Y.; Yu, Q.; Wei, Q.; Luo, X.; Xie, Y.; Adeli, E.; Wang, Y.; et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. *Med. Image Anal.* **2024**, *97*, 103280. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Isensee, F.; Jaeger, P.F.; Kohl, S.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Zhao, Q.; Huang, Y.; Wang, B.; Zhang, Q.; Antwi-Afari, M.F.; Zhao, W.; Sun, B. An integrated deep learning framework for CT-based mesoscopic segmentation and quantitative analysis of 3D-printed concrete. *J. Build. Eng.* **2026**, *123*, 115869. [\[CrossRef\]](#)
13. Werner, B.; Rauchs, G.; Havlásek, P.; Sýkora, J.; Belouettar, S.; Patzák, B. Efficient AI-aided segmentation of concrete CT scans using a predictive annotation interface. *KSCE J. Civ. Eng.* **2026**, *30*, 100416. [\[CrossRef\]](#)
14. Zhang, Z.; Zhu, E.; Li, J.; Wang, B.; Cai, W.; Zhu, C. Meso-structure segmentation of concrete CT image using lightweight DeepLabv3 plus with attention mechanisms: X-ray computed tomography and mesoscopic modeling. *J. Build. Eng.* **2026**, *121*, 115694. [\[CrossRef\]](#)
15. Carrara, P.; Kruse, R.; Bentz, D.P.; Lunardelli, M.; Leusmann, T.; Varady, P.A.; De Lorenzis, L. Improved mesoscale segmentation of concrete from 3D X-ray images using contrast enhancers. *Cem. Concr. Compos.* **2018**, *93*, 30–42. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Thakur, M.M.; Enright, S.; Hurley, R.C. Phase segmentation in X-ray CT images of concrete with implications for mesoscale modeling. *Constr. Build. Mater.* **2023**, *403*, 133033. [\[CrossRef\]](#)
17. Yang, S.; Mao, L.; Zhou, M.; Chang, X.; Lui, J.; Li, D.; Bai, J. A fine-segmentation algorithm for XCT images of multiphase composite building materials based on deep learning. *J. Build. Eng.* **2024**, *97*, 110918. [\[CrossRef\]](#)
18. Xie, L.; Zhou, B.; Yao, Y.; Xu, Q.; Zhong, R.; Zhou, H.; Wang, J. Image segmentation-based quantification of phase distribution in coarse aggregate ultra-high performance concrete and its impact on mechanical properties. *Cem. Concr. Compos.* **2026**, *165*, 106326. [\[CrossRef\]](#)
19. Chen, D.; Li, Y.; Tao, J.; Li, Y.; Zhang, S.; Shan, X.; Wang, T.; Qiao, Z.; Zhao, R.; Fan, X.; et al. Deep learning-based segmentation model for permeable concrete meso-structures. *Comput.-Aided Civ. Infrastruct. Eng.* **2024**, *39*, 3626–3645. [\[CrossRef\]](#)
20. Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment anything in medical images. *Nat. Commun.* **2024**, *15*, 654. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Shi, W.; Zhang, P.; Li, Y.; Jiang, Z. Segment anything model for few-shot medical image segmentation with domain tuning. *Complex Intell. Syst.* **2025**, *11*, 37. [\[CrossRef\]](#)
22. Kim, H.-T.; Park, K. Computed tomography (CT) image-based analysis of concrete microstructure using virtual element method. *Compos. Struct.* **2022**, *299*, 115937. [\[CrossRef\]](#)
23. Tabernik, D.; Šuc, M.; Skočaj, D. Automated detection and segmentation of cracks in concrete surfaces using joined segmentation and classification deep neural network. *Constr. Build. Mater.* **2023**, *408*, 133582. [\[CrossRef\]](#)
24. Dong, Y.; Su, C.; Qiao, P.; Sun, L. Microstructural crack segmentation of three-dimensional concrete images based on deep convolutional neural networks. *Constr. Build. Mater.* **2020**, *253*, 119185. [\[CrossRef\]](#)
25. Barisin, T.; Jung, C.; Müsebeck, F.; Redenbach, C.; Schladitz, K. Methods for segmenting cracks in 3d images of concrete: A comparison based on semi-synthetic images. *Pattern. Recogn.* **2022**, *129*, 108747. [\[CrossRef\]](#)
26. Baduge, S.K.; Thilakarathna, S.; Perera, J.S.; Ruwanpathirana, G.P.; Doyle, L.; Duckett, M.; Lee, J.; Saenda, J.; Mendis, P. Assessment of crack severity of asphalt pavements using deep learning algorithms and geospatial system. *Constr. Build. Mater.* **2023**, *401*, 132684. [\[CrossRef\]](#)
27. Dang, L.M.; Wang, H.; Li, Y.; Nguyen, L.Q.; Nguyen, T.N.; Song, H.K.; Moon, H. Lightweight pixel-level semantic segmentation and analysis for sewer defects using deep learning. *Constr. Build. Mater.* **2023**, *371*, 130792. [\[CrossRef\]](#)
28. Nian, T.; Xue, S.; Ge, J.; Han, Z.; An, Z. Application of U-Net++ deep learning network for segmentation and processing of asphalt mixture X-ray CT images. *Appl. Soft Comput.* **2025**, *177*, 113308. [\[CrossRef\]](#)

29. Zhang, W.; Tian, C.; Yang, J.; Xiao, P.; Wu, Z.; Li, H.; Sun, W. Vision and deep learning-based tunnel surface defect detection: Technological evolution, challenges, and prospects. *Measurement* **2026**, *257*, 118853. [[CrossRef](#)]
30. Lin, Z.; Dall'Ara, E.; Guo, L. A novel mean shape based post-processing method for enhancing deep learning lower-limb muscle segmentation accuracy. *PLoS ONE* **2024**, *19*, e0308664. [[CrossRef](#)] [[PubMed](#)]
31. Wu, Y.; Wang, Y.; Li, D.; Zhang, J. Two-step detection of concrete internal condition using array ultrasound and deep learning. *NDT E Int.* **2023**, *139*, 102945. [[CrossRef](#)]
32. Qian, H. Deep Learning-Based Concrete Image Analysis and Generalization Capability. Ph.D. Thesis, Nanyang Technological University, Singapore, 2024.
33. *DL/T 5330-2015*; Design Specification for Mix Proportion of Hydraulic Concrete. National Energy Administration: Beijing, China, 2015.

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.