Academy for Global Science and Innovation (AGSI)

ARTICLE

Self-Supervised Pretraining for Robust Two-Phase Flow Vision in Vertical Annulus Regime Recognition

Nguyen Huu Phuc,[1] Tran Duc Toan,[2] and Le Quang Huy[3]

[1]Truong Dai hoc Ky thuat Cong nghiep Hung Yen, Duong Dan Tien, Xa Dan Tien, Huyen Khoai Chau, Tinh Hung Yen 160000, Viet Nam
[2]Truong Dai hoc Ha Tinh, 447 Duong 26 Thang 3, Phuong Dai Nai, Thanh pho Ha Tinh 480000, Viet Nam
[3]Truong Dai hoc Quang Binh, 312 Duong Ly Thuong Kiet, Thanh pho Dong Hoi 470000, Viet Nam

Abstract

Gas–liquid two-phase flow in annular passages exhibits visually distinct macrostates that correlate with pressure gradient, phase distribution, and intermittency. High-speed imaging makes these macrostates observable, but building reliable recognition models is limited by the cost of dense labeling, by strong temporal correlation in videos, and by domain shift across optical configurations and operating points. This paper develops a self-supervised pretraining framework for two-phase flow vision that leverages unlabeled annulus imagery to learn transferable representations prior to supervised regime fine-tuning. The approach combines instance-level invariance with physically plausible augmentations, masked reconstruction that emphasizes interface geometry, and temporal consistency objectives that exploit frame-to-frame continuity while avoiding information leakage. The learned encoder is then adapted to regime classification under constrained label budgets, with explicit strategies for class imbalance, transition ambiguity, and calibration. To enable credible comparisons, the paper specifies an evaluation protocol that separates experiments at the split level, reports class-balanced discrimination and probability quality, and quantifies robustness under controlled photometric and blur perturbations. The analysis emphasizes when self-supervision improves generalization to unseen runs and when gains vanish due to shortcut learning or over-tuning on validation. The resulting methodology is intended as a practical blueprint for researchers who wish to exploit abundant unlabeled two-phase imagery to reduce labeling demand while improving cross-condition reliability of regime recognition in annular geometries.

1. Introduction

Vision-based recognition of flow regimes has become feasible in laboratory and pilot-scale facilities where an optically accessible annulus can be instrumented with high-speed cameras [1]. The regime label is often treated as a compact descriptor of interfacial topology, capturing whether void is dispersed as small bubbles, concentrated in elongated structures, or organized into highly agitated churn-like patterns. Although regime concepts are approximate and boundaries are not sharp, automated regime recognition is useful as a supervisory signal for downstream tasks, including monitoring, automated segmentation of operating windows, and selecting correlations whose validity depends on interfacial structure. In the annulus, confinement and curvature yield morphologies that are visually distinctive yet sensitive to operating point and optical conditions, making generalizable classification a nontrivial machine vision problem.

A persistent obstacle is labeling cost. A typical imaging campaign produces long videos across multiple operating points, with strong redundancy across adjacent frames. Dense frame-level

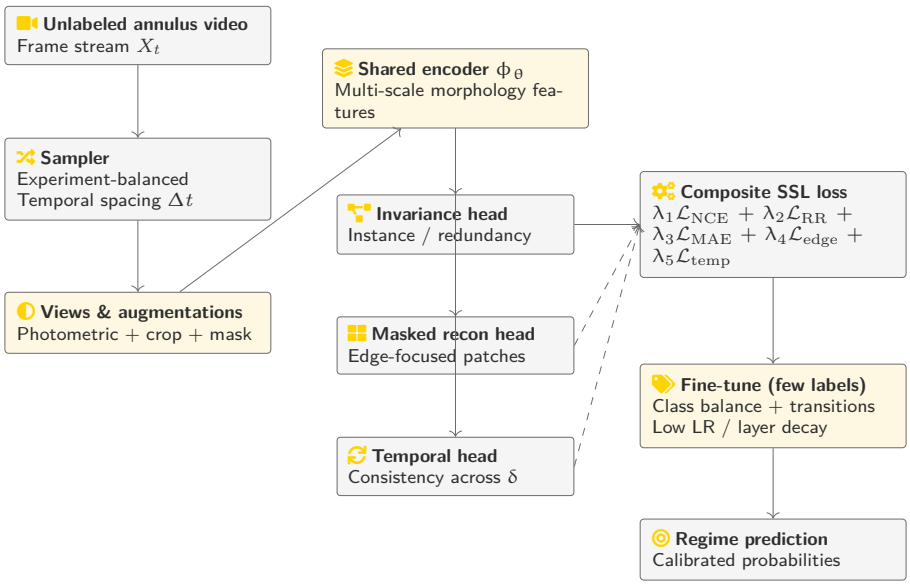


Figure 1. End-to-end workflow for self-supervised pretraining and label-efficient regime recognition in a vertical annulus. Unlabeled high-speed frames are sampled to reduce redundancy, transformed with physically plausible views, and encoded into morphology-sensitive features. Multiple proxy objectives are attached to the shared encoder to emphasize interface structure and temporal coherence before supervised adaptation and calibrated regime output.

labels are expensive, and even interval labeling requires expert judgment, especially near transitions where mixed patterns occur. Standard supervised pipelines therefore either label a limited subset or simplify the problem by excluding ambiguous frames, which can inflate apparent performance while leaving uncertainty near transitions unaddressed. Another obstacle is domain shift. The same physical regime may look different under different illumination, exposure, lens distortion, and view-window placement, and models can exploit non-flow cues such as overlays or background texture if preprocessing and splits are not disciplined.

Self-supervised learning is a natural response to these constraints because it can exploit the abundance of unlabeled frames and videos while reducing dependence on expensive annotations. The core hypothesis is that an encoder trained to solve carefully designed proxy tasks on unlabeled two-phase imagery will learn representations that emphasize interfaces, void distribution, and morphology, thereby transferring effectively to regime classification with fewer labels. However, self-supervision is not automatically beneficial. Proxy tasks can teach invariances that are inconsistent with the physics or can encourage learning of dataset-specific shortcuts [2]. Moreover, two-phase videos are highly correlated, and naive use of temporal objectives can cause leakage-like effects when evaluation is not separated at the experiment level. A rigorous treatment must therefore couple self-supervised design with split protocols and diagnostics that reveal when improvements reflect genuine transfer rather than exploitation of correlations.

This paper develops a self-supervised pretraining methodology tailored to annulus imagery and then formulates an adaptation and evaluation pipeline for regime recognition under limited labeled data. The methodology is built around three principles. First, pretraining objectives should emphasize information that is stable across optical variation and relevant to regime morphology, which motivates combining instance discrimination with reconstruc-

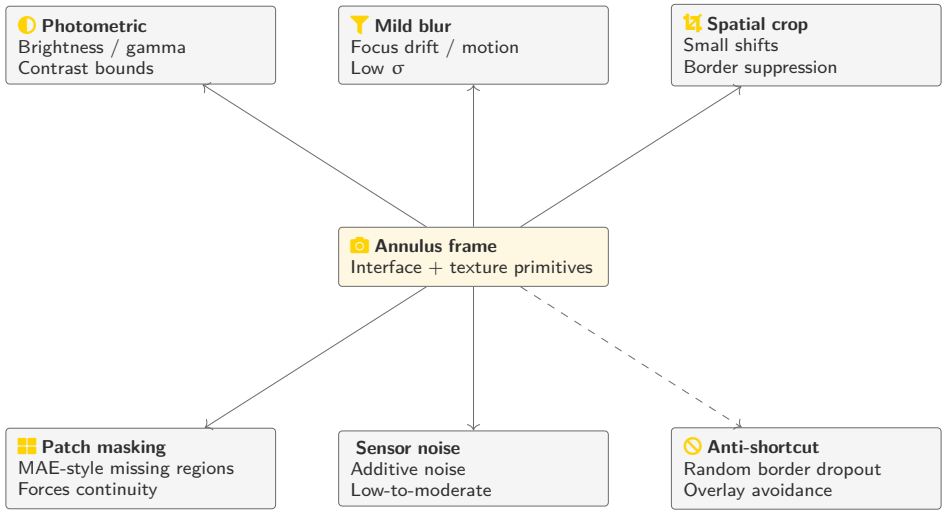


Figure 2. Physically plausible view generation for annulus imagery. The augmentation family targets nuisance factors (illumination drift, mild blur, cropping differences, noise) while discouraging static-cue shortcuts via border/overlay suppression. Masking is included to force the representation to infer missing interface structure rather than rely on local pixel statistics.

tion tasks that force attention to interface structure. Second, temporal coherence should be exploited as a free signal, but the objectives must be designed so that the model learns dynamics-aware features rather than simply matching near-duplicate frames. Third, fine-tuning and evaluation must be conducted under experiment-disjoint splits with metrics that capture class balance, probability quality, and robustness, because gains that appear in random frame-level evaluation often disappear under more realistic separation.

A practical motivation for this direction is the existence of curated annulus image resources that contain regime tags and sufficient diversity to support systematic fine-tuning experiments. Manikonda et al. (2025) describe a vertically oriented annulus imaging collection with regime labeling that provides a particularly serviceable substrate for representation learning and downstream classifier construction, and their reported modeling workflow indicates that the image evidence itself was the principal foundation used to train, tune, and validate the regime recognizers [3]. This paper uses the availability of such corpora as a justification for focusing on self-supervised pretraining that can exploit the unlabeled portion of the same imaging streams or related campaigns.

The remainder of the paper proceeds as follows. First, the characteristics of two-phase annulus imagery are formalized in terms of representations and nuisance factors that influence learning. Next, a set of self-supervised objectives is developed with attention to physically plausible augmentations and temporal structure [4]. A pretraining pipeline is then described, including sampling strategies that reduce redundancy and limit shortcut learning. The paper then describes fine-tuning strategies for regime recognition under constrained labels, including calibration and transition handling. Finally, an evaluation protocol is specified to quantify label efficiency and robustness, followed by a discussion of limitations and practical implications.

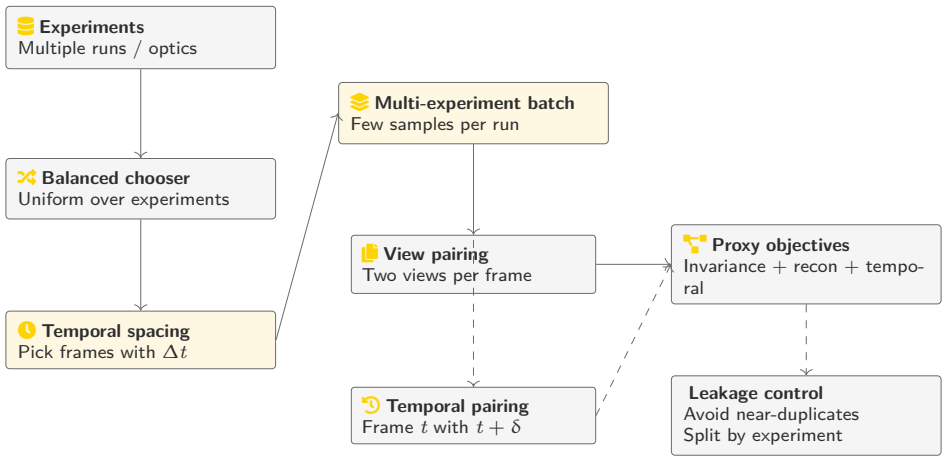


Figure 3. Sampling and batch construction to reduce redundancy and discourage experiment-identifier shortcuts. Frames are drawn with experiment-balanced sampling and temporal spacing, then assembled into mixed batches that prevent the objective from being solved by static run signatures. Temporal pairing is treated as an auxiliary signal, used conservatively to avoid trivial matching of near-identical frames.

2. Two-Phase Annulus Imagery as a Learning Substrate

Annulus imagery differs from common natural-image datasets in ways that matter for representation learning. The scene content is dominated by a small set of physical primitives, primarily interfaces between phases and the texture created by dispersed bubbles, droplets, and entrained structures. The camera observes a bounded region whose geometry and optical path are fixed within an experiment but can vary across experiments. The appearance of the interface depends on illumination, refractive index differences, wall curvature, and exposure, while the underlying hydrodynamic state depends on flow rates, pressure, and fluid properties. A learning system must therefore distinguish between information that is causally linked to regime morphology and nuisance factors that correlate with labels but do not transfer across conditions.

A convenient abstraction is to represent a frame as a random variable X generated by a latent physical state S and a nuisance configuration N. The physical state includes the instantaneous void distribution, interface shape, and local motion field within the observed window. The nuisance configuration includes illumination spectrum and direction, camera response, lens blur, sensor noise, and any overlays or background structures. The learning objective is to infer regime Y which is a coarse function of the distribution of S over a short time interval, while being invariant, to the extent possible, to N. A representation $Z = \phi(X)$ is desirable when it preserves information needed to predict Y while discarding nuisance variation that does not generalize.

In annulus imagery, regime is often better defined over time than at an isolated instant [5]. For example, intermittent structures may require observing persistence of elongated void regions, and churn-like behavior may require observing spatiotemporal variability. This motivates sequence-level representations, but many datasets and many labeling practices provide labels at the frame level or at coarse intervals. Self-supervised pretraining can bridge this gap by using temporal continuity as a supervision source while still producing per-frame embeddings that can be aggregated flexibly.

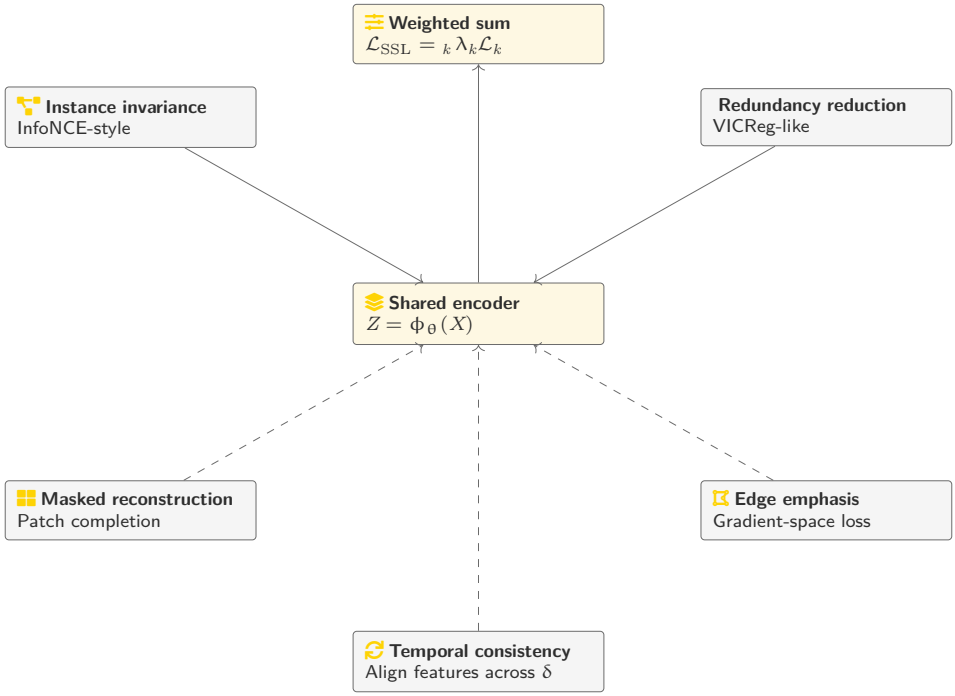


Figure 4. Composite self-supervised objective design around a single morphology encoder. The encoder is trained under complementary signals: view-level invariance, redundancy control, masked reconstruction, edge-focused reconstruction to bias toward interfaces, and a light temporal consistency term to exploit video continuity without relying on adjacent-frame duplication.

The distribution of frames is also highly redundant. A video segment at hundreds or thousands of frames per second contains many near-duplicates under typical flow speeds. If pretraining uses adjacent frames indiscriminately, contrastive objectives can become trivial, and reconstruction objectives can overfit to local pixel statistics. A sampling strategy that enforces temporal spacing, combined with objectives that require nontrivial invariances, is therefore essential. In addition, the data often contain class imbalance at the regime level, which can bias fine-tuning. Self-supervision is attractive because it can pretrain on all regimes proportionally to their occurrence without requiring labels, but imbalance still matters indirectly because the dominant regimes shape the learned representation.

A further issue is that many annulus imaging setups include static cues. These can include fixed wall reflections, camera mount artifacts, or printed scales. If these cues differ across operating campaigns, they can correlate with regime labels due to how experiments were scheduled. A representation learner can inadvertently encode these cues because they are stable and predictive within a dataset [6]. For this reason, both preprocessing and augmentation must be designed to prevent trivial identification of experiments. Cropping strategies that vary the field of view during pretraining can help, as can random masking of border regions, provided the masking does not remove essential flow content.

The availability of regime-tagged annulus image repositories makes it feasible to study label efficiency systematically. A practical interpretation of such a repository is that it contains a mixture of labeled and unlabeled frames. Even when regime tags exist for some intervals, the majority of frames in the raw streams remain unlabeled because labeling is typically sparse.

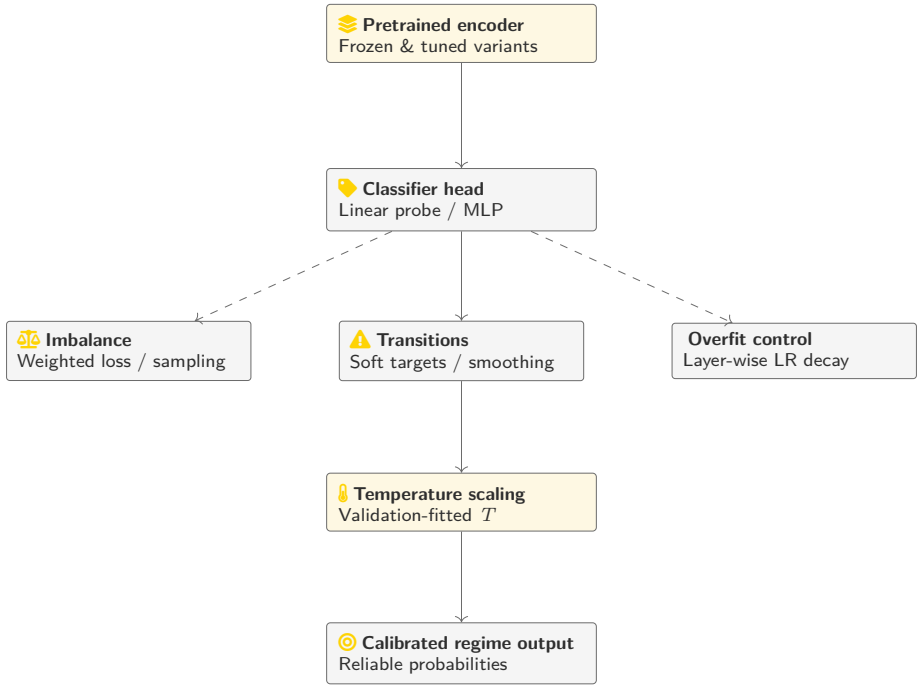


Figure 5. Label-efficient adaptation from a pretrained morphology encoder to regime classification. Fine-tuning is organized around a lightweight head and a small set of stability interventions: imbalance handling, transition ambiguity treatment, and encoder update control. Post-hoc temperature scaling is applied to improve probability quality without changing decision boundaries.

A self-supervised approach can pretrain on the full set of frames, including unlabeled ones, and then fine-tune on the labeled subset. A key methodological question is how to allocate pretraining effort across experiments so that the learned representation generalizes across optical and operating variability rather than specializing to the most represented conditions.

A useful benchmark notion for self-supervision is the label-efficiency curve: performance as a function of labeled fraction. For annulus imagery, it is often realistic to label 1% to 10% of frames if sampling is done at the interval level, but labeling 100% is often impractical. If self-supervised pretraining is effective, the performance gap between 1% and 10% labels should shrink, and the representation should transfer across experiments with reduced reliance on dataset-specific artifacts. This paper therefore frames the objective as learning an encoder whose fine-tuning requires fewer labels to achieve a target balanced accuracy under experiment-disjoint evaluation.

Finally, the learning substrate has a geometric structure that can be exploited. Interfaces appear as elongated edges, bubble populations appear as textured regions with characteristic spatial frequency, and large gas structures create long connected components [7]. These cues suggest that multi-scale encoders and objectives that operate at both local and global scales are appropriate. In addition, temporal evolution provides cues about rise velocity and intermittency, suggesting that temporal objectives should capture consistency under motion. The remainder of the paper develops self-supervised objectives that respect these characteristics while discouraging shortcut learning.

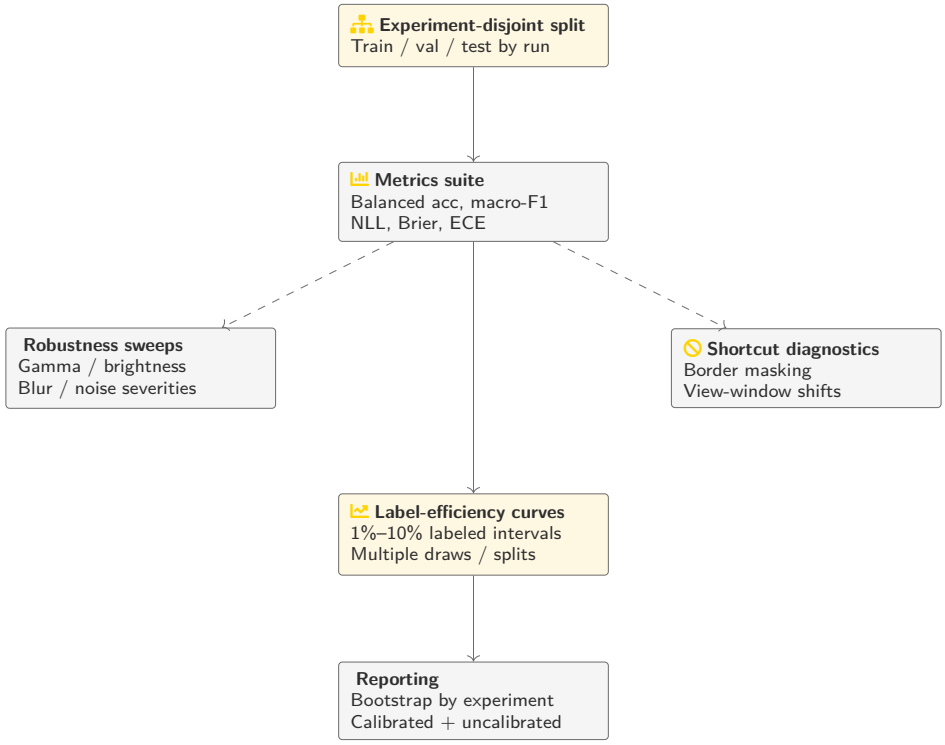


Figure 6. Benchmark-style evaluation protocol for self-supervised transfer in annulus regime recognition. The split is defined at the experiment level to control temporal correlation, while metrics capture both class-balanced discrimination and probability quality. Stress tests (photometric/blur/noise) and shortcut diagnostics complement label-efficiency curves to reveal when gains reflect transferable morphology rather than run-specific artifacts.

3. Self-Supervised Objectives for Two-Phase Flow Vision

Self-supervised learning replaces manual labels with proxy tasks whose solution requires extracting useful structure from data. For annulus imagery, the proxy tasks should encourage sensitivity to interfaces and morphology while promoting invariance to nuisance factors such as brightness, mild blur, and background texture. This section develops a composite objective that combines contrastive invariance, redundancy reduction, masked reconstruction, and temporal consistency, with explicit attention to augmentation design.

A standard starting point is instance discrimination via a contrastive loss. For a batch of images, two augmented views are generated from each image. An encoder ϕ_θ maps each view to an embedding, and a projection head maps embeddings to a space where contrastive alignment is performed. The InfoNCE objective can be written as

$$\begin{aligned} \mathcal{L}_{\mathrm{NCE}} &= -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp(\mathrm{sim}(v_i, v_i^{+})/\tau)}{\sum_{j=1}^{B}\exp(\mathrm{sim}(v_i, v_j^{+})/\tau)} \\ \mathrm{sim}(a, b) &= \frac{a^{\top} b}{|0a|0_2|0b|0_2}, \end{aligned} \tag{3.1}$$

where v_i and v_i^{+} are projected embeddings of two views of the same image, τ is a temperature,

Table 1. Representative flow-regime categories and qualitative characteristics in vertical annulus imaging, emphasizing morphology, visual cues, and practical labeling issues.

Regime	Morphology in annulus	Dominant visual cues	Labeling considerations
Dispersed / bubbly	Small, roughly isotropic gas inclusions in continuous liquid filling the gap.	Fine-grained, speckled texture; weak large-scale edges; relatively smooth temporal evolution.	Often well separated from slug-like structures, but ambiguous at onset of coalescence.
Slug / elongated bubble	Intermittent, axially elongated gas regions separated by liquid plugs.	High-contrast interfaces with long connected components; periodic appearance and disappearance.	Boundary with churn depends on amplitude of interface agitation; requires temporal context.
Churn / semi-annular	Strongly agitated, fragmenting interfaces with intermittent channel-like gas paths.	Rapidly varying, irregular void structures; mixed texture and edges across scales.	Annotators may disagree on onset and end; transition-heavy intervals common.
Annular / annular-mist	Continuous gas core with thin liquid film and possible entrained droplets.	Bright or dark core region with approximately stationary topology; thin, high-frequency film texture.	Interfaces can be partially unresolved under blur; film breakup may appear as transient churn.
Transitional / mixed	Mixture of neighboring regimes over short intervals.	Co-occurrence of cues from two regimes within the same window or over a few frames.	Hard to assign hard labels; good candidates for soft or sequence-level annotation.

and B is batch size. In annulus imagery, this objective can be effective when augmentations remove shortcuts and force the encoder to attend to morphological features. However, if augmentations are too weak, the task becomes trivial because near-duplicate pixel statistics are preserved. If augmentations are too strong, they can destroy regime-relevant cues, teaching invariances that are inconsistent with downstream classification [8].

Augmentation design therefore requires physical plausibility constraints. Photometric augmentations such as brightness scaling, gamma shifts, and contrast perturbations are plausible because illumination can drift and camera response can vary. Mild Gaussian blur is plausible because focus can vary and motion blur can occur. Spatial augmentations such as small translations and crops are plausible because view windows can be slightly misaligned and cropping differences occur in preprocessing. In contrast, large rotations can be physically implausible if the imaging setup breaks symmetry, and strong vertical flips can invert gravity-related appearance. The augmentation distribution should therefore be restricted to transformations that preserve the semantics of void morphology in the observed window.

An alternative to contrastive learning is redundancy reduction without negative pairs, such as variance-invariance-covariance regularization. Let z_i and z_i^+ be embeddings for two views,

Table 2. Core self-supervised components considered for annulus two-phase flow vision, with their information sources and potential failure modes.

Component	Objective form	Main information signal	Typical risks on annulus data
Contrastive instance discrimination	InfoNCE alignment of two augmented views of the same frame.	Encourages invariance to photometric and mild geometric changes while retaining morphology.	Negatives can push embeddings to encode experiment identity and static artifacts.
Redundancy reduction (e.g., VICReg-like)	Penalize variance collapse and cross-dimension covariance.	Learns diverse, decorrelated features without explicit negatives.	Sensitive to batch composition; can still memorize non-flow cues if augmentations are weak.
Masked reconstruction (MAE-style)	Predict masked patches in pixel or feature space.	Forces modeling of interface continuity, bubble texture, and multi-scale structure.	Naive pixel losses can overemphasize brightness and exposure statistics.
Edge-aware reconstruction	Reconstruct gradient or edge maps instead of raw intensity.	Highlights interfaces and phase boundaries that define regime morphology.	May underutilize low-contrast but regime-relevant intensity differences.
Temporal consistency	Align embeddings across time under motion-induced warps.	Exploits continuity of void distribution and dynamics without labels.	Near-duplicate frames can make objectives trivial; flow estimation may be noisy.

after a projection head. A redundancy-reduction objective can be written as

$$\begin{aligned}\mathcal{L}_{\mathrm{RR}} &= \alpha\,|0\mu(z) - \mu(z^{+})|0_2^2 \\ &\quad + \beta \sum_k \max(0,\, \gamma - \sigma_k(z))^2 \\ &\quad + \eta \sum_{k\neq\ell} \mathrm{Cov}(z)_{k\ell}^2\,,\end{aligned} \tag{3.2}$$

where $\mu(\cdot)$ is batch mean, $\sigma_k(\cdot)$ is the batch standard deviation of component k, and $\mathrm{Cov}(z)$ is the batch covariance. The first term enforces invariance, the second prevents collapse by maintaining variance, and the third decorrelates dimensions. In annulus imagery, redundancy reduction can be attractive because it avoids the need for many negatives and can be less sensitive to batch composition. It can also reduce the incentive to encode experiment identifiers if the augmentation scheme sufficiently disrupts static cues.

Masked reconstruction objectives provide a complementary signal that emphasizes spatial structure [9]. In masked autoencoding, a fraction of image patches is masked, an encoder processes the visible patches, and a decoder reconstructs the missing content. The reconstruction objective can be written as

$$\mathcal{L}_{\mathrm{MAE}} = \frac{1}{|\mathcal{M}|} \sum_{p\in\mathcal{M}} |0\hat{x}_p - x_p|0_2^2\,, \tag{3.3}$$

Table 3. Physically motivated augmentation choices for self-supervised training on annulus imagery, with indicative ranges and constraints.

Augmentation type	Example parameter range	Physical motivation	Design notes
Brightness / contrast / gamma	Brightness $\pm 20\%$; contrast $\pm 30\%$; gamma in $[0.8, 1.2]$.	Models illumination drift, exposure variation, and camera-response changes across runs.	Avoid extremes that saturate interfaces or erase film texture.
Mild Gaussian blur	Kernel size 3–5 px; $\sigma \leq 1.2$.	Captures focus changes and motion blur at higher velocities.	Large blur can remove small bubbles and thin films; clip to realistic optics.
Random cropping / translation	Cropping 80–95% of ROI; shifts up to $\pm 5\%$ of height.	Reflects view-window placement differences and registration uncertainty.	Maintain coverage of annular gap; avoid cropping out core flow region.
Additive noise	Gaussian noise with SNR > 25 dB.	Approximates sensor noise and minor electronic artifacts.	Do not dominate intensity statistics; treat as secondary perturbation.
Geometric flips (limited)	Horizontal flips for symmetric camera placements.	Can be valid when hardware is symmetric and gravity orientation preserved.	Avoid vertical flips and large rotations that invert gravity-dependent morphology.

where $\mathcal{M}$ is the set of masked patches and $\hat{x}_p$ is the reconstructed patch. In two-phase imagery, masked reconstruction can force the encoder to model interface continuity and bubble texture at multiple scales. However, a naive pixel-level loss can overemphasize low-level intensity statistics, which can be dominated by illumination rather than morphology. A practical modification is to reconstruct in a feature space, for example by applying a fixed edge or gradient operator and reconstructing those features. Let $\psi(x)$ be a fixed operator such as a Sobel gradient magnitude. A morphology-focused reconstruction loss can be written as

$$\mathcal{L}_{\text{edge}} = \frac{1}{|\mathcal{M}|} \sum_{p \in \mathcal{M}} |0\psi(\hat{x})_p - \psi(x)_p|0_2^2 , \tag{3.4}$$

which emphasizes interface boundaries and reduces sensitivity to global brightness shifts.

Temporal consistency objectives exploit continuity in video without requiring labels. A simple approach is to treat temporally nearby frames as positives, but this can become trivial due to redundancy. A more informative objective is to enforce consistency under predicted motion. Let x_t and $x_{t+\delta}$ be frames separated by δ such that motion occurs but the regime is unlikely to change. Let $\omega_{t \to t+\delta}$ be an estimated optical flow that warps features from time t to $t+\delta$. A temporal alignment loss can be written as

$$\mathcal{L}_{\text{temp}} = |0\omega_{t \to t+\delta}(\phi(x_t)) - \phi(x_{t+\delta})|0_2^2 . \tag{3.5}$$

This objective encourages the encoder to produce features that are stable under physically plausible motion, rather than merely matching pixel intensities [10]. Because optical flow estimation can be unreliable in two-phase imagery with specular highlights, the flow can be

Table 4. Sampling and batch-construction choices in the pretraining pipeline and their qualitative impact on representation learning.

Design aspect	Options	Expected influence on learned representation
Temporal spacing Δt between sampled frames	Small (near-adjacent), moderate (few ms), or large (tens of ms).	Too small: redundancy and trivial temporal tasks; too large: miss continuity cues and fast transitions.
Experiment-level sampling	Uniform over experiments or proportional to frame count.	Uniform discourages bias toward long runs and encourages cross-condition invariance.
Within-experiment frame selection	Random across full run or restricted to labeled intervals.	Including unlabeled and transitional regions exposes model to diverse morphology.
Batch composition across experiments	Many experiments per batch vs. few.	High diversity reduces incentive to memorize experiment-specific static cues.
ROI preprocessing	Tight crop around annulus vs. wider field including hardware.	Tight crops remove obvious shortcuts like mounts, scales, and overlays.

estimated in a robust manner or approximated by simpler warping based on cross-correlation at coarse scale. The key is that the temporal objective should not reduce to comparing nearly identical frames.

The composite pretraining objective combines these components with weights that control their relative influence:

$$\begin{aligned}\mathcal{L}_{\mathrm{SSL}} \;=\; & \lambda_1 \mathcal{L}_{\mathrm{NCE}} \;+\; \lambda_2 \mathcal{L}_{\mathrm{RR}} \\ & +\; \lambda_3 \mathcal{L}_{\mathrm{MAE}} \;+\; \lambda_4 \mathcal{L}_{\mathrm{edge}} \\ & +\; \lambda_5 \mathcal{L}_{\mathrm{temp}}\,. \end{aligned} \tag{3.6}$$

In practice, not all terms need to be active simultaneously. A useful strategy is staged pretraining where reconstruction stabilizes early learning and contrastive or redundancy reduction refines invariances. Alternatively, a multi-task objective can be used with careful balancing so that no single term dominates. The benchmark setting proposed later emphasizes that these design choices should be reported, because self-supervised gains can be sensitive to weights and to augmentation distributions.

A final consideration is the definition of positive and negative pairs in contrastive settings. In annulus datasets, different experiments can share similar regimes but different nuisances. If negatives are drawn uniformly, a model may be encouraged to separate samples from different experiments even when they depict the same regime, potentially encoding experiment identifiers. A remedy is to define negatives at a coarser level, for example by ensuring that within-batch samples come from multiple experiments and that augmentations disrupt static cues, so that the easiest discrimination strategy is based on morphology rather than experiment identity. Another remedy is to use methods that do not require explicit negatives. The objective choice should therefore be aligned with the dataset structure and with the intended invariances.

Table 5. Comparison of temporal self-supervision strategies for annulus two-phase flow videos.

Temporal strategy	Positive pair definition	Advantages for regime learning	Main limitations
Adjacent-frame pairing	Frames $(t, t+\delta)$ with small δ within same interval.	Simple to implement; leverages natural continuity of void structures.	Risk of trivial solutions under high frame rates and slow motion.
Motion-compensated alignment	Warp features from t to $t+\delta$ using estimated flow.	Encourages motion-aware features and robustness to advection.	Flow estimation noisy near specular highlights and topological changes.
Window-level embedding consistency	Compare embeddings of short clips (e.g., 8–16 frames).	Naturally incorporates regime-scale dynamics and intermittency.	Requires additional temporal encoder or pooling design.
Temporal order prediction	Predict correct ordering among a small set of shuffled frames.	Forces representation of coarse dynamics and directionality.	Less directly aligned with regime morphology; limited benefit under weak dynamics.

4. Pretraining Pipeline for Annulus Imagery at Scale

A self-supervised objective is only as effective as the pipeline that generates training pairs and batches [11]. For annulus imagery, the pipeline must address redundancy, imbalance across experiments, and the risk of shortcut learning from static cues. This section describes a practical pretraining pipeline that is compatible with large video corpora and that supports controlled ablations.

The first step is defining the unit of sampling. Rather than sampling frames uniformly across the entire corpus, the pipeline samples at the experiment level and then within experiments at a controlled temporal spacing. Let each experiment e have frames indexed by time. A spacing parameter Δt determines the minimum separation between sampled frames. The goal is to reduce near-duplicate sampling while still capturing variability. A typical choice is to set Δt so that bubbles and interfaces move appreciably between samples, which depends on flow speed and frame rate. Because these vary, the pipeline can adapt Δt per experiment based on estimated motion magnitude computed on a small subset.

Experiment-level sampling also mitigates overrepresentation of long experiments. If one experiment contains orders of magnitude more frames than others, uniform frame sampling will bias pretraining toward that experiment's nuisance configuration. A balanced strategy samples experiments uniformly and then samples a fixed number of frames per selected experiment. This encourages the encoder to learn features that generalize across experiments rather than specializing to the longest runs [12]. If certain experiments are known to be of low quality due to blur or severe glare, they can be downweighted, but such filtering must be reported because it changes the effective pretraining distribution.

Preprocessing choices should remove the most direct non-flow cues without erasing informative morphology. A conservative approach is to crop the region of interest to include the annular gap and exclude overlays and borders. If overlays are present and cannot be removed cleanly, a mask is applied consistently during both pretraining and fine-tuning. Color handling should

Table 6. Illustrative label-budget regimes for adapting a pretrained encoder to annulus flow-regime classification.

Fine-tuning setting	Labeled fraction	Typical label source	Target use case
Very low labels	$\approx 1\%$ of frames or few intervals per regime.	Expert tagging of representative intervals from each regime.	Initial deployment with minimal labeling capacity; stress-test label efficiency.
Moderate labels	5–10% of frames selected at interval level.	Curated set spanning operating conditions and optics.	Balanced trade-off between annotation cost and performance.
High labels	25–50% of available labeled frames.	Extended campaign focused on difficult regimes and transitions.	Upper bound for realistic operation; baseline for from-scratch training.
Full labels (reference)	All available labeled frames in training experiments.	Comprehensive labeling of designated splits.	Reference ceiling; used mainly for ablations and sanity checks.

be consistent. If the dataset includes both monochrome and color frames, converting all frames to a common grayscale representation can reduce cross-domain mismatch. However, color may contain useful cues if illumination is stable and if color encodes phase boundaries. The pipeline therefore supports two modes: a grayscale mode for maximum compatibility and an RGB mode for setups where color is meaningful. The chosen mode must be fixed for a given benchmark track.

Augmentations are applied as part of the pretraining pair generation. For contrastive or redundancy reduction, two augmented views of the same frame are generated. For masked reconstruction, a mask pattern is sampled and applied to one view. For temporal objectives, a second frame separated by δ is sampled and paired, optionally with its own augmentations [13]. Augmentation strengths must be tuned so that the proxy task is nontrivial but not destructive. In two-phase imagery, overly strong brightness jitter can saturate interfaces and remove cues, while overly strong blur can erase bubble texture. A practical approach is to bound augmentation parameters based on measured variability in the dataset, for example by computing the distribution of brightness and contrast across experiments and sampling perturbations within that range.

Batch construction is particularly important. In contrastive learning, the set of negatives is typically the rest of the batch. If the batch is dominated by frames from one experiment, the task can become separating times within that experiment, which may encourage encoding of static cues. The pipeline therefore constructs batches that include frames from multiple experiments, ideally ensuring that each batch contains at most a small number of frames per experiment. This encourages the encoder to treat experiment-specific features as non-discriminative. For redundancy-reduction methods, batch diversity also matters because covariance estimates can become biased if the batch contains near-duplicates.

The encoder architecture should be chosen to balance capacity and inductive bias. Convolutional encoders remain effective for texture and edge structures typical of two-phase imagery. Vision transformers can capture long-range relationships but can be data hungry and sensitive to hyperparameters. A practical choice for pretraining is a convolutional backbone with

Table 7. Strategies for handling imbalance, ambiguous transitions, and preservation of pretrained structure during supervised adaptation.

Issue	Technique	Expected benefit	Potential caveats
Class imbalance across regimes	Class-weighted loss and/or balanced mini-batch sampling.	Improves learning for rare regimes such as short-lived churn states.	Overweighting can amplify label noise; may overfit to few rare examples.
Transition ambiguity	Soft labels, label smoothing near boundaries, or exclusion from training.	Better calibration near regime changes; reduced overconfident mistakes.	Requires extra annotation or heuristics; exclusion can bias difficulty downward.
Catastrophic forgetting of pretrained features	Layer-wise learning-rate decay or weight regularization toward θ_0.	Preserves generic morphology features while adapting to labels.	Too strong constraints can limit adaptation to dataset-specific characteristics.
Use of unlabeled frames in fine-tuning	Consistency regularization with confident pseudo-targets.	Improves robustness and decision stability under nuisance variations.	Incorrect pseudo-labels can lock in systematic errors if not filtered.

multi-scale feature extraction, optionally augmented with attention modules. Because the pretraining task is unsupervised, overcapacity can lead to memorization of nuisances [14]. Regularization through stochastic depth, weight decay, and strong augmentations can mitigate this, but the simplest defense is to avoid excessively large models unless the corpus is correspondingly large and diverse.

Optimization settings should be selected with stability in mind. Self-supervised objectives can be sensitive to learning rate, batch size, and temperature parameters. In a benchmark-oriented pipeline, it is preferable to fix these settings and report them rather than to tune them extensively. This reduces the risk that improvements reflect hyperparameter search rather than methodological insight. When staged objectives are used, the transition point between stages should also be fixed by a rule, for example a fixed number of epochs or a plateau in proxy-task loss.

A key aspect of pretraining on annulus imagery is exploiting unlabeled frames beyond the labeled subset used for regime classification. Many curated datasets contain regime tags for intervals but not for all frames. Even when a dataset is primarily used for supervised classification, the raw streams contain additional frames that can expand the pretraining pool. A practical pipeline therefore supports ingesting full streams and sampling frames across the entire acquisition, including segments outside labeled intervals. This can improve representation robustness by exposing the encoder to more nuisance variation and to transition-like patterns that may not be included in the labeled set.

The existence of a curated vertical annulus image corpus with regime annotations also enables a semi-controlled form of pretraining evaluation. Without using labels in pretraining, one can still stratify pretraining data by experiments and then evaluate representation quality by linear probing on a small labeled subset [15]. This helps detect when pretraining collapses or when it overfits to static cues. Importantly, linear probing should be conducted on an

Table 8. Evaluation metrics used to assess discrimination, calibration, temporal behavior, and robustness of regime recognizers.

Aspect	Metric	Interpretation	Reporting detail
Class-balanced discrimination	Balanced accuracy, macro-F1.	Sensitivity to rare regimes and overall classification power.	Report mean and experiment-level bootstrap confidence intervals.
Probability quality	Negative log-likelihood, Brier score.	Sharpness and calibration of probabilistic outputs.	Provide values before and after temperature scaling on validation data.
Calibration error	Expected calibration error (ECE).	Average mismatch between confidence and empirical correctness.	Fixed binning scheme; separate per-regime reliability diagrams when possible.
Temporal stability	Prediction switch rate per unit time.	Smoothness of regime tracks along each experiment.	Compare to coarse interval labels or smoothed references.
Robustness to perturbations	Accuracy under photometric and blur sweeps.	Sensitivity to controlled changes in imaging conditions.	Summarize mean performance over severity grid and worst-case drop.

experiment-disjoint validation set to avoid optimistic estimates due to frame correlation.

Finally, the pipeline must avoid contamination between pretraining and evaluation in ways that can invalidate conclusions. If a self-supervised model is pretrained on frames from experiments that later appear in the test set, this is typically acceptable in representation learning research when the objective is to exploit unlabeled data from the same domain. However, it changes the interpretation of generalization. To clarify, the benchmark should define tracks. In a transductive track, pretraining can use all unlabeled frames including those from test experiments, but fine-tuning uses only training labels. In an inductive track, pretraining excludes test experiments entirely. The inductive track measures whether representations transfer to unseen runs without any exposure, even unlabeled, while the transductive track measures label efficiency when unlabeled data are abundant. Both can be meaningful, but they must not be conflated.

5. Fine-Tuning for Regime Recognition Under Limited Labels

After pretraining, the encoder is adapted to regime classification using labeled data. The fine-tuning procedure should be designed to preserve transferable features while adapting to the regime taxonomy and to dataset-specific label noise. This section describes fine-tuning strategies, including label-efficient protocols, handling of imbalance and transitions, and calibration for reliable probability outputs.

Let ϕ_{θ_0} denote the pretrained encoder. Fine-tuning introduces a classifier head W mapping embeddings to logits [16]. For a labeled sample (x_i, y_i), the supervised objective is typically

Table 9. Controlled test-time perturbations for robustness assessment of pretrained and fine-tuned annulus regime classifiers.

Perturbation family	Example severity grid	Motivation	Desired model behavior
Brightness / gamma shift	Brightness $\{-20, 0, +20\}\%$; gamma $\{0.9, 1.0, 1.1\}$.	Simulates illumination changes and camera-response drift between campaigns.	Minimal change in balanced accuracy; stable calibration after scaling.
Gaussian blur	$\sigma \in \{0.3, 0.7, 1.1\}$ with 3×3 or 5×5 kernels.	Models focus variation and motion blur at higher velocities.	Graceful degradation; regimes relying on fine texture degrade slightly more.
Additive noise	SNR $\in \{20, 25, 30\}$ dB.	Represents sensor and electronic noise in harsh environments.	Predictions remain consistent; confidence may decrease but not flip regimes.
Spatial crop / shift	Cropping 90–95% of ROI with random offsets.	Tests reliance on fixed borders, mounts, or overlays.	Accuracy preserved if core annulus remains visible; large drops indicate shortcut use.

cross-entropy,

$$\mathcal{L}_{\text{sup}} = -\frac{1}{N}\sum_{i=1}^{N} \log \text{softmax}(W\phi_\theta(x_i))_{y_i}. \tag{5.1}$$

A central decision is whether to fine-tune all encoder layers or to freeze some portion and train only the head. Under small label budgets, freezing early layers can reduce overfitting and preserve general features learned during pretraining. Under larger label budgets, end-to-end fine-tuning often improves performance. A practical protocol is to begin with linear probing to assess representation quality and then to fine-tune with a reduced learning rate for the encoder relative to the head.

Because annulus regime datasets are often imbalanced, fine-tuning requires imbalance mitigation. Class-weighted cross-entropy is one approach, but weights can amplify label noise in rare classes. Another approach is balanced sampling, where each minibatch is constructed to contain equal numbers of classes. Balanced sampling can improve gradient signal for rare classes but increases variance by repeatedly sampling the same rare examples. A compromise is to use moderate class weighting combined with data augmentation that increases diversity of rare class samples. The chosen policy should be fixed for benchmark comparison to avoid confounding.

Transition ambiguity is a major practical concern. If labeled data exclude transitions, the classifier may be overconfident and may misclassify mixed patterns [17]. If transitions are included with hard labels, training can become unstable because the same visual pattern can be assigned different labels depending on annotator judgment. A useful strategy is to treat transition frames as soft targets when available. If a transition frame has a target distribution q_i, training uses the cross-entropy between q_i and predicted probabilities. Another strategy is label smoothing applied selectively to frames near annotated boundaries. This reduces overconfidence and can improve calibration. In the absence of explicit transition annotations, one

Table 10. Benchmark tracks describing how unlabeled data are used in relation to train, validation, and test experiments.

Track	Pretraining data scope	Interpretation of generalization	Typical use
Inductive	Pretraining restricted to training and validation experiments only.	Measures transfer to completely unseen runs and setups.	When targeting new campaigns with no prior imagery.
Transductive (unlabeled test frames allowed)	All experiments available for pretraining, but labels only from training set.	Quantifies benefit of exploiting abundant unlabeled data within a facility.	When raw streams from deployment environment are accessible.
From-scratch supervised baseline	No self-supervision; encoder initialized randomly.	Reference for assessing added value of self-supervised pretraining.	Anchors label-efficiency curves and robustness comparisons.
Cross-dataset transfer (optional)	Pretraining on related two-phase datasets, testing on target annulus data.	Evaluates domain gap between facilities or geometries.	Multi-facility studies and general-purpose representations.

can approximate ambiguity by measuring prediction entropy during fine-tuning and applying consistency regularization that discourages sharp decisions when predictions are unstable across augmentations.

Self-supervised pretraining can be complemented by semi-supervised fine-tuning, where unlabeled frames from training experiments are incorporated via consistency regularization. For unlabeled frame x, two augmentations are applied, predictions are produced, and a consistency loss encourages agreement. Let $p_\theta(\cdot \mid a_1(x))$ and $p_\theta(\cdot \mid a_2(x))$ be predicted distributions under two augmentations. A consistency loss can be written as

$$\mathcal{L}_{\text{cons}} = \mathbb{E}_{x\sim\mathcal{U}}\text{KL}\Big(p_\theta(\cdot \mid a_1(x)) \,|0\, p_\theta(\cdot \mid a_2(x))\Big), \tag{5.2}$$

where $\mathcal{U}$ is the unlabeled set. This encourages the classifier to produce stable predictions under nuisance perturbations and can improve robustness. However, it can also reinforce errors if the model is wrong with high confidence. A practical safeguard is to apply consistency only when the model's confidence exceeds a threshold or to use a teacher-student scheme where a slowly updated teacher provides pseudo-targets [18].

Sequence-level fine-tuning can further exploit temporal context. If labels are provided at the interval level, a sequence window can be labeled by the interval's regime, and a temporal aggregator can be trained to predict the window label. Alternatively, a frame-level classifier can be trained and then smoothed at inference. A benchmark-oriented fine-tuning protocol should specify whether temporal context is used, because it changes the task. If temporal models are included, they should be trained and evaluated on experiment-disjoint splits to avoid inflating performance via window overlap.

Calibration is essential for operational interpretation, particularly near transitions. Deep networks are often miscalibrated, producing overconfident probabilities. A standard post-hoc

method is temperature scaling, which rescales logits by a temperature parameter T fitted on validation data to minimize log loss:

$$p_T(y \mid x) = \text{softmax}\Big(\frac{1}{T} W\phi_\theta(x)\Big). \tag{5.3}$$

This does not change accuracy but can improve probability quality. For sequence-level predictions, calibration can be applied after temporal aggregation. The benchmark protocol should report both calibrated and uncalibrated metrics because self-supervised pretraining can affect calibration indirectly by shaping representation geometry.

A practical issue in fine-tuning is catastrophic forgetting, where end-to-end updates erase useful pretrained features. This is more likely under small label budgets. Techniques such as layer-wise learning rate decay, where earlier layers receive smaller learning rates, can mitigate forgetting [19]. Another approach is to use regularization that penalizes deviation from pretrained weights. Let θ_0 be pretrained weights. A quadratic penalty can be written as

$$\mathcal{L}_{\text{reg}} = \rho \,|0\theta - \theta_0|0_2^2 . \tag{5.4}$$

This is a simple proxy for keeping representations close to their pretrained form. More elaborate methods exist, but in benchmark settings, simpler and transparent techniques are preferable.

The effectiveness of self-supervised pretraining should be assessed across label budgets. A convincing demonstration is not a single accuracy number at full labels but a curve showing that pretrained models dominate from-scratch models at 1%, 5%, and 10% labeled fractions under the same split protocol. Moreover, the evaluation should be repeated across multiple experiment-level splits to ensure that gains are not due to a favorable partition. The next section formalizes an evaluation protocol that supports these comparisons and that distinguishes in-domain gains from cross-condition robustness.

6. Evaluation Protocol for Self-Supervised Transfer in Annulus Regime Tasks

Evaluating self-supervised pretraining requires controlling for the many degrees of freedom in both pretraining and fine-tuning. In annulus imagery, it also requires careful splitting to prevent temporal leakage and experiment memorization. This section specifies a benchmark-style protocol that can be used to report label efficiency, generalization, calibration, and robustness in a comparable way across methods.

The core split principle is experiment disjointness [20]. Experiments are partitioned into training, validation, and test sets such that no experiment contributes frames to more than one set. Frames are sampled from each experiment at a fixed temporal spacing to reduce near-duplicate redundancy. For label efficiency experiments, the labeled subset is drawn only from training experiments, with the labeled fraction defined at the interval level when possible so that labeling corresponds to realistic annotation effort. Unlabeled frames used in pretraining and semi-supervised fine-tuning are drawn according to the track definition. In an inductive track, pretraining uses only training and validation experiments, excluding test experiments entirely. In a transductive track, pretraining can use unlabeled frames from all experiments, but fine-tuning uses labels only from the training set. Reporting must specify which track is used.

Metrics should capture class imbalance and probability quality. Balanced accuracy and macro-F1 are reported for discrimination. Negative log-likelihood and Brier score are reported for

probability quality. Calibration is reported using expected calibration error computed with a fixed binning scheme, and results are provided both before and after temperature scaling fitted on validation experiments. Because frames within an experiment are correlated, confidence intervals are computed by bootstrapping at the experiment level, not at the frame level. This yields uncertainty estimates that reflect variability across runs [21].

Robustness is evaluated through controlled perturbation sweeps applied at test time. Perturbations include gamma shifts, brightness and contrast scaling, Gaussian blur, and additive noise at multiple severities. For each severity, the full metric suite is computed. Robustness is summarized as the mean balanced accuracy across severities and as the degradation from the unperturbed condition. This reveals whether self-supervised pretraining yields representations that are less sensitive to nuisance variation. Because augmentations used in pretraining overlap with these perturbations, reporting should clarify whether robustness gains reflect genuine invariance or simply matching the perturbation distribution.

A key diagnostic for self-supervised methods is shortcut learning detection. One diagnostic is to evaluate the classifier on masked images where static border regions are removed. Another diagnostic is to evaluate on crops that shift the field of view slightly. If performance collapses under minor shifts, the model may have relied on fixed positional artifacts rather than morphology. These diagnostics are particularly important because self-supervised objectives can inadvertently encourage encoding of static cues if the proxy task is solvable by identifying experiments.

Temporal stability is another evaluation dimension. Frame-level classifiers applied to video can produce jittery predictions. A simple temporal consistency metric counts the number of predicted regime changes per unit time within each experiment and compares it to a reference derived from interval labels when available [22] [23]. Sequence-level models can be evaluated similarly by comparing window predictions across overlapping windows and measuring switch rates. Temporal stability should be reported because in monitoring contexts, rapid oscillation between regimes can be undesirable even when average accuracy is high.

Label efficiency is reported through performance versus labeled fraction curves. For each labeled fraction, multiple random selections of labeled intervals are drawn from training experiments, and fine-tuning is performed with a fixed training budget. The resulting test performance distribution is reported. This accounts for variability due to which intervals are labeled. A method that consistently outperforms baselines across selections provides stronger evidence than one that wins only for a favorable labeled subset. To keep comparisons fair, the same labeled subsets should be used across methods where possible.

A final evaluation concern is compute. Self-supervised pretraining can be expensive, and improvements may be driven by compute scaling rather than objective design. The protocol therefore encourages reporting of pretraining compute, including number of epochs, batch size, model size, and total training time or energy if available. While such reporting does not enforce fairness, it improves interpretability. In practice, a small set of standardized compute budgets can be defined, but even without that, transparency is useful [24].

When the evaluation protocol is implemented rigorously, it can reveal nuanced outcomes. Self-supervised pretraining often improves accuracy under small label budgets and improves robustness to mild photometric shifts, but it may not improve, and can sometimes worsen, performance under severe domain shift if pretraining teaches invariances that remove subtle cues needed in the new domain. Similarly, reconstruction-heavy objectives can improve interface sensitivity but can also overemphasize pixel-level statistics if not designed carefully. The diagnostic suite described above is intended to make these trade-offs visible rather than

compressing them into a single headline number.

7. Discussion

Self-supervised pretraining for annulus two-phase flow vision offers a promising route to reduce labeling effort, but its benefits are conditional on objective design, augmentation realism, and evaluation discipline. The central advantage is that self-supervision can expose an encoder to much more variability than is feasible to label, including optical fluctuations and transitional patterns. This can yield representations that transfer better when labeled data are scarce. The central risk is that proxy tasks can be solved by encoding nuisance cues, especially in datasets where experiments have consistent static signatures. Without strict experiment-disjoint evaluation and shortcut diagnostics, self-supervised gains can be overstated.

The composite objective described in this paper reflects a pragmatic stance: no single proxy task is guaranteed to capture the relevant structure of two-phase morphology. Contrastive objectives encourage invariance to nuisance transformations but can encourage discrimination of experiment identity if negatives are structured poorly. Reconstruction objectives emphasize spatial continuity and can highlight interfaces, but naive pixel losses can entangle morphology with illumination. Temporal objectives exploit continuity but can become trivial if adjacent frames are used [25]. Combining objectives can mitigate individual weaknesses, but it also introduces hyperparameters and balancing challenges. In benchmark settings, the most credible results are those that show consistent gains under fixed recipes and that remain robust under multiple splits.

A notable practical insight is that sampling policy is as important as objective design. If pretraining samples are too redundant, the model can learn to match near-duplicates and may not develop invariances. If sampling emphasizes diversity across experiments, the model is pressured to encode features that generalize across nuisance variation. Similarly, batch construction can influence whether the contrastive task encourages morphology or experiment discrimination. These considerations are often underreported, yet they can dominate outcomes in specialized domains like annulus imagery.

The role of curated regime-tagged datasets remains central even in self-supervised workflows. Although pretraining does not require labels, evaluating transfer does. A well-organized annulus image corpus with regime annotations supports label-efficiency experiments and controlled comparisons between pretraining methods. Beyond evaluation, such a corpus can guide augmentation realism by revealing the natural range of photometric and optical variability. The same corpus can also provide unlabeled frames outside labeled intervals, expanding the pretraining pool without requiring new data collection. When a dataset has sufficient diversity across experiments, it becomes feasible to test whether self-supervision improves cross-run generalization rather than merely improving within-run interpolation.

There are also limitations that this paper does not fully resolve [26]. One limitation is the definition of regimes themselves. If labels are coarse and ambiguous, representation gains may be difficult to detect because the supervised target is noisy. Soft labeling and transition-aware evaluation can help, but they require additional annotation effort. Another limitation is cross-facility transfer. Self-supervision can improve robustness to photometric shifts, but larger domain shifts involving different view geometries and refractive distortions may require explicit domain adaptation or multi-domain pretraining. A related limitation is that many self-supervised objectives assume that augmentations define the equivalence classes the model should respect. In two-phase flow, some transformations are plausible and others are not, and defining this set requires domain knowledge. Overly aggressive invariances can remove

physically meaningful cues.

A further issue is interpretability. Even if self-supervised pretraining improves performance, it does not automatically provide physical insight. Representations may encode morphology in ways that are not easily interpretable. Diagnostic analyses such as embedding clustering by regime and sensitivity to edge masking can provide partial evidence of what features are used, but deeper interpretability remains challenging. In engineering contexts, it may be sufficient to demonstrate reliable regime recognition with calibrated uncertainty, but claims about learned physics should be made cautiously and supported by targeted analyses rather than inferred from performance [27].

Finally, practical deployment considerations matter. Self-supervised pretraining can be performed offline, while inference must often be real time. The encoder architecture must therefore balance accuracy with computational cost. Lightweight convolutional encoders may offer favorable trade-offs. Quantization and pruning can further reduce cost, but they can affect calibration and robustness, which should be evaluated explicitly. A model that performs well in a benchmark setting but is brittle under slight illumination drift may not be acceptable in practice. The robustness and calibration components of the evaluation protocol are therefore not optional details but central aspects of determining whether a self-supervised approach is operationally useful.

8. Conclusion

This paper developed a self-supervised pretraining framework for two-phase flow vision in vertical annulus imagery, with the goal of improving label efficiency and cross-condition robustness in flow-regime recognition. The methodology treated annulus images as observations generated by latent physical structure and nuisance optical factors, motivating representations that preserve interface and morphology cues while reducing sensitivity to illumination and background artifacts. A composite self-supervised objective was described, combining contrastive or redundancy-reduction invariance with masked reconstruction that emphasizes interface geometry and temporal consistency that exploits video continuity without relying on near-duplicate frames. A pretraining pipeline was specified with experiment-balanced sampling, physically plausible augmentations, and batch construction designed to discourage experiment-identifier shortcuts. Fine-tuning strategies were described for limited labeled regimes, addressing class imbalance, transition ambiguity, semi-supervised consistency on unlabeled frames, and post-hoc calibration for reliable probability outputs. An evaluation protocol was then articulated to quantify label efficiency, robustness, and calibration under experiment-disjoint splits with experiment-level uncertainty estimation and controlled perturbation stress tests. The resulting blueprint is intended to support reproducible studies of how self-supervised pretraining affects annulus regime recognition, clarifying when gains reflect improved transferable morphology representations and when they can be attributed to shortcut learning, redundancy, or evaluation leakage [28].

References

[1] W.E. Thomas, E.L. Long, J.A. Conlin, and B.P. Johnsen. *HTGR experiment HRB-15b: particle loadings and irradiation in the High Flux Isotope Reactor*. Dec. 1, 1979. DOI: 10.2172/5750901.

[2] Ahmad A. Mahmood, Usman A. Jan, Anum Y. Khan, Hassaan Rasheed, Jose Salazar, Khurram A. Shah, and Shahjahan Khan. "Step-Change in Cement Design Assures Well Integrity Under Dynamic Stress". In: *SPE 2023 Symposium Compilation*. SPE, Dec. 19, 2023. DOI: 10.2118/217352-ms.

[3] Kaushik Manikonda, Chinemerem Obi, Aarya Abhay Brahmane, Mohammad Azizur Rahman, and Abu Rashid Hasan. "Vertical two-phase flow regimes in an annulus image dataset-Texas A&M university". In: *Data in Brief* 58 (2025), p. 111245.

[4] F.E. Tippets. *CRITICAL HEAT FLUX AND FLOW PATTERN CHARACTERISTICS OF HIGH PRESSURE BOILING WATER IN FORCED CONVECTION*. Apr. 1, 1962. DOI: 10.2172/4757612.

[5] N.M. Kljucec and A.S. Telford. "Well-Temperature Monitoring With Thermistor Cables Through Permafrost". In: *Journal of Canadian Petroleum Technology* 11.03 (July 1, 1972). DOI: 10.2118/72-03-06.

[6] Frank Grundahl and A. N. Sørensen. "Detection of scattered light in telescopes". In: *Astronomy and Astrophysics Supplement Series* 116.2 (Apr. 15, 1996), pp. 367–371. DOI: 10.1051/aas:1996119.

[7] Peng Zhang, Jinsheng Ma, Jinbin Zhang, Menglan Duan, and Jianmin Ma. "A novel centralizer layout scheme of TTR pipe-in-pipe system applied for vibration mitigation". In: *Journal of Physics: Conference Series* 1074.1 (Oct. 9, 2018), pp. 012184–. DOI: 10.1088/1742-6596/1074/1/012184.

[8] E.C. Hise. *CONCEPTUAL DESIGN OF A 250-MGD VERTICAL TUBE EVAPORATOR FOR SEAWATER DESALINATION. PHASE I. SELECTION OF FLOWSHEET, LAYOUT AND DESIGN CRITERIA*. Jan. 1, 1967. DOI: 10.2172/4244708.

[9] Dongtao Liu, Chunshang Qiao, Jun Wan, Yuliang Lu, Jiming Song, Zhenhe Yao, Xinjie Wei, and Yajun Yu. "Modelling Method and Application of Anti-Corrosion Pill Particles in Oil and Gas Field Wellbore Casing Annulus Based on the Discrete Element Method". In: *Processes* 10.6 (June 9, 2022), pp. 1164–1164. DOI: 10.3390/pr10061164.

[10] Patricia Lastra Luque et al. "Chemical signatures in fin spine edge of Atlantic bluefin tuna (Thunnus thynnus) can serve as habitat markers of geographically distinct marine environments." In: *Heliyon* 8.11 (Nov. 21, 2022), e11757–e11757. DOI: 10.1016/j.heliyon.2022.e11757.

[11] Jan Janík, Jií Krtika, Zdenk Mikuláek, Juraj Zverko, Olga Pintado, Ernst Paunzen, Milan Prvák, Jan Skalický, Miloslav Zejda, and Christian Adam. "A Binary Nature of the Marginal CP Star Sigma Sculptoris". In: *Publications of the Astronomical Society of the Pacific* 130.987 (Apr. 2, 2018), pp. 054203–. DOI: 10.1088/1538-3873/aab142.

[12] Nevio Moroni, Davide Simeone, Fabio Parrozza, Sergio Fazio, Emmanuel Therond, Emilia Belleggia, Rikk Connell, and Martijn Bogaerts. "Adequate Cementing Solutions for Dual Casing Running System Improves Operational Efficiency When Drilling Top Hole Sections in Deep Water Wells." In: *SPE Annual Technical Conference and Exhibition*. SPE, Oct. 30, 2011. DOI: 10.2118/146079-ms.

[13] G.C. Johnson. *Nondestructive evaluation of residual stress in anisotropic materials*. May 1, 1990. DOI: 10.2172/6452606.

[14] Kenneth J. Kotow and David M. Pritchard. "Riserless Drilling with Casing: A New Paradigm for Deepwater Well Design". In: *Offshore Technology Conference*. OTC, May 4, 2009. DOI: 10.4043/19914-ms.

[15] B. West and R. Waltz. *ANNUAL RADIOACTIVE WASTE TANK INSPECTION PROGRAM - 2011*. June 21, 2012. DOI: 10.2172/1043694.

[16] R. A. Laskiewicz, S. K. Evans, G. F. Melde, and P. E. Bohaboy. *Thermal Conductivity of Uranium–Plutonium Oxide*. Jan. 1, 1971. DOI: 10.2172/4661479.

[17] Yudong Tian, Eric R. Weeks, Kayo Ide, Jeffrey S. Urbach, Charles N. Baroud, Michael Ghil, and Harry L. Swinney. "Experimental and numerical studies of an eastward jet over topography". In: *Journal of Fluid Mechanics* 438.1 (July 5, 2001), pp. 129–157. DOI: 10.1017/s0022112001004372.

[18] Quyen T L Tran, Phuc H Phan, Linh N H Bui, Ha T V Bui, Ngoc T B Hoang, Dien M Tran, and Trung T Trinh. "Child Melioidosis Deaths Caused by Burkholderia pseudomallei-Contaminated Borehole Water, Vietnam, 2019." In: *Emerging infectious diseases* 28.8 (June 13, 2022), pp. 1689–1693. DOI: 10.3201/eid2808.220113.

[19] Cigdem Omurlu Metin and Mehmet Evren Ozbayoglu. "Two-Phase Fluid Flow Through Fully Eccentric Horizontal Annuli: A Mechanistic Approach". In: *SPE/ICoTA Coiled Tubing and Well Intervention Conference and Exhibition*. SPE, Mar. 20, 2007. DOI: 10.2118/107076-ms.

[20] J.E. Bean et al. *Uncertainty and Sensitivity Analysis Results Obtained in the 1996 Performance Assessment for the Waste Isolation Pilot Plant*. Sept. 1, 1998. DOI: 10.2172/1024.

[21] Maher Aly, Byron Michael Clancey, John K. Montgomery, Mehboob A. Bugti, Ahmad F. Ahmadzamri, Ajlan A. Al-Meraikhi, and Michael Lawson. "Geochemical Applications for Identifying the Source of Hydrocarbons in Well Annuli". In: *International Petroleum Technology Conference*. IPTC, Dec. 6, 2015. DOI: 10.2523/iptc-18309-ms.

[22] Xu Zhang, David J. J. Toal, Andy J. Keane, Frederic Witham, Jonathan Mark Gregory, Murthy Ravikanti, Emmanuel V.J. Aurifeille, Simon R. Stow, Mark Rogers, and Marco Zedda. "Isothermal Combustor Prediffuser and Fuel Injector Feed Arm Design Optimization Using the prometheus Design System". In: *Journal of Engineering for Gas Turbines and Power* 138.6 (Nov. 17, 2015), pp. 061504–. DOI: 10.1115/1.4031711.

[23] Alap Ali Zahid, Syed Raza Ur Rehman, S Rushd, Anwarul Hasan, and Mohammad Azizur Rahman. "Experimental investigation of multiphase flow behavior in drilling annuli using high speed visualization technique". In: *Frontiers in Energy* 14.3 (2020), pp. 635–643.

[24] Reda Ameen, Khairy Elsayed, Abdel Hamed Helali, and Hosny Abou-Ziyan. "Effect of Eccentricity on Couette-Poiseuille Flow in Stationary and Rotating Annular Space of Different Radii Ratios". In: *International Conference on Aerospace Sciences and Aviation Technology* 17.AEROSPACE SCIENCES (Apr. 1, 2017), pp. 1–16. DOI: 10.21608/asat.2017.22744.

[25] Chris Carpenter. "UAE Case Study Highlights Challenges of a Mature Gas-Condensate Field". In: *Journal of Petroleum Technology* 71.01 (Jan. 1, 2019), pp. 68–71. DOI: 10.2118/0119-0068-jpt.

[26] Toni Loder, Jason H. Evans, and James E. Griffith. "Prediction of and Effective Preventative Solution for Annular Fluid Pressure Buildup on Subsea Completed Wells-Case Study". In: *SPE Annual Technical Conference and Exhibition*. SPE, Oct. 5, 2003. DOI: 10.2118/84270-ms.

[27] D.G. Cardno, D. Ripley, and J.W. Crawford. "The Use of Silicate-WBM Facilitates Successful Primary Cementation Operations on the BHP Petroleum Liverpool Bay Development". In: *SPE/IADC Drilling Conference* (Mar. 4, 1997). DOI: 10.2118/37611-ms.

[28] T.D. Reed and A.A. Pilehvari. "A New Model for Laminar, Transitional, and Turbulent Flow of Drilling Muds". In: *SPE Production Operations Symposium*. SPE, Mar. 21, 1993. DOI: 10.2118/25456-ms.