PHYSICS-AWARE HEAT PARTITIONING NETWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal measurements are critical for understanding two-phase boiling heat transfer, yet experimentally resolving phase-specific heat flux contributions remains expensive and challenging. Building on well-established heat-partitioning formulas, we introduce PAPA-Net, a Physics-Aware Heat Partitioning Network for optical-only, mechanism-resolved wall heat flux prediction. PAPA-Net factorizes each heat-transfer component into a soft physical support mask and a nonnegative heat-amplitude field, yielding dry-area, microlayer, and single-phase contributions that sum to the total prediction. Sparse teacher masks are used only during training to assign mechanism identity, while inference requires only optical observations. Physics-coupled objectives enforce consistency between local predictions, mechanism contributions, and integrated heat balance. Across training, validation and external zero-shot conditions, PAPA-Net maintains stable, noncollapsed physical partitions and, on the held-out condition, reduces integrated heat error by 20.4% relative to the next-best optical-only method. Ablations, label-efficiency, robustness, uncertainty, and controlled synthetic experiments further show that the mask-amplitude factorization yields mechanism-aligned heat flux partitions rather than a cosmetic latent decomposition.

1 INTRODUCTION

Boiling is one of the most efficient heat transfer mechanisms, yet its wall-level physics remains difficult to measure directly. High-speed optical imaging can capture bubble growth, dry spots, contact line motion, and liquid-vapor interface dynamics, but the quantities that determine heat transfer performance, such as local wall heat flux and its per-mechanism decomposition, are usually hidden or expensive to obtain. Recent AI models for boiling and phase-change systems have made progress in forecasting multiphase dynamics, reconstructing thermal fields, and predicting engineering quantities such as critical heat flux (Suh et al., 2024; Hassan et al., 2023; 2025; Fu et al., 2026; Khodakarami et al., 2025; Furlong et al., 2025). However, most learned models predict a monolithic temperature or heat flux field. They do not explain how much of the heat transfer arises from dry area effects, μ-layer activity, or single-phase convection, which are precisely the mechanisms used in classical heat flux partitioning theory.

The difficulty is that mechanism-resolved heat partitioning is underconstrained by total heat flux supervision alone. A standard regressor can predict a plausible $\hat{q}''$ while assigning nearly all heat to one latent component, duplicating mechanisms across components, or producing components with no physical correspondence. Conversely, fully guided models that require externally provided phase masks at test time can recover clean partitions, whereas PAPA-Net takes only the optical image at inference and predicts the mechanism supports internally, without mask preprocessing. The central question is therefore: can sparse teacher masks be used during training to align predicted components with physical heat flux mechanisms, while requiring only optical observations at inference?

The same issue appears broadly in scientific machine learning and representation learning: a dense field predictor can be accurate while its internal decomposition is physically meaningless, echoing known failures of unsupervised disentanglement, shortcut learning, and models that are right for the wrong reasons (Locatello et al., 2019; Geirhos et al., 2020; Ross et al., 2017; Margeloiu et al., 2021). For boiling, this failure is especially important because the total heat flux is not the only object of interest; engineers also need to know which mechanism produced it. Thus, the ML problem is not simply dense regression, but weakly supervised physical disentanglement under an optical-only test time constraint. A useful model must therefore satisfy three requirements simultaneously: predict

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the heat flux map, assign heat to named physical mechanisms, and avoid using privileged phase masks at inference.

We propose PAPA-Net, a Physics-Aware Heat Partitioning Network for mechanism-resolved wall heat flux prediction using only the optical image as model input at inference. PAPA-Net factorizes each heat transfer component as a product of a soft physical support mask and a nonnegative heat amplitude field, $\hat{q}''_k = \hat{z}_k \hat{a}_k$, so that the total prediction remains $\hat{q}'' = \sum_k \hat{q}''_k$ while each component retains a physical interpretation. Sparse teacher masks are used only during training to assign the identity of the mechanism and prevent collapse; at inference, the model requires only optical input.

Our contributions are:

- We formulate optical-only mechanism-resolved heat flux partitioning as a physical-science ML problem for phase change heat transfer, where sparse teacher masks are available only during training, and no phase annotations are used at inference.
- We introduce PAPA-Net, a mask-amplitude factorization that separates physical support $\hat{z}_k$ from mechanism-specific heat intensity $\hat{a}_k$, yielding interpretable dry area, μ-layer, and single phase (liquid phase) heat flux components.
- We show that PAPA-Net recovers noncollapsed, mechanism-aligned partitions on IID and held-out boiling conditions, outperforming unguided decomposition in physical component quality while remaining competitive with strong optical-only baselines in total heat flux prediction.
- We provide extensive validation through architecture and loss ablations, label efficiency studies, robustness tests, physics-aware heat flux metrics, runtime analysis, external zero-shot transfer to a second boiling dataset under changed thermal and acquisition conditions, and a controlled synthetic partition benchmark.

2 Related Work

AI for boiling and phase-change heat transfer. Recent work has established boiling and phase-change heat transfer as a demanding frontier for scientific machine learning, where sharp liquid-vapor interfaces, stochastic nucleation, topology changes, and near-wall thermal gradients make standard surrogate modeling difficult. Broad reviews have argued that AI can help extract hidden physical quantities from high-speed imaging and phase change data streams (Suh et al., 2024), while datasets and benchmarks such as BubbleML and BubbleML 2.0 have enabled systematic learning of multiphase boiling dynamics (Hassan et al., 2023; 2025). Closely related efforts have used learning to analyze bubble geometry and motion (Chang et al., 2023; Seong et al., 2023), forecast boiling evolution (Hassan et al., 2025), infer temperature fields from optical observations (Fu et al., 2026; Khodakarami et al., 2025), and model broader multiphase flows with neural operators or large-scale scientific machine learning (SciML) benchmarks (Wen et al., 2022; Shadkhah et al., 2025). Other work targets engineering quantities such as critical heat flux using hybrid physics ML models (Furlong et al., 2025). These advances show that learned models can recover or forecast difficult thermal fluid fields, but they usually predict total fields, temperature maps, future states, or scalar heat transfer quantities. In contrast, PAPA-Net targets *mechanism-resolved heat flux partitioning*: it predicts the total wall heat flux map while decomposing it into dry area, μ-layer, and single phase components to account for contributions by the existence of vapor, evaporation, and liquid convection, using only optical input at inference.

Physics-aware inverse field reconstruction. A broader line of SciML work learns mappings between partial observations, boundary conditions, and dense physical fields using physics-informed learning, neural operators, and generative inverse models (Raissi et al., 2019; Karniadakis et al., 2021; Lu et al., 2021; Li et al., 2021; Kovachki et al., 2023). Recent diffusion and flow-based models further frame PDE reconstruction as conditional generation under partial observability, enabling uncertainty-aware recovery of hidden fields from sparse or incomplete measurements (Li et al., 2024; Huang et al., 2024; Jacobsen et al., 2025; Bastek et al., 2025; Haitsiukevich et al., 2024). These approaches are powerful for reconstructing temperature, velocity, or full PDE states, but they typically output monolithic fields rather than physically named heat transfer mechanisms. PAPA-Net is complementary: instead of only inferring a hidden dense field, it imposes a teacher-guided mask-amplitude factorization so that the predicted heat flux is decomposed into interpretable phase-specific contributions.

Heat flux partitioning and mechanistic boiling models. Mechanistic boiling theory has long decomposed wall heat transfer into physically distinct pathways, including microlayer evaporation, quenching, single-phase convective exchange, dry area, and vapor-related mechanisms (Rohsenow, 2022; Mikic & Rohsenow, 1969; Judd & Hwang, 1976; Del Valle & Kenning, 1985; Kurul & Podowski, 1990; Richenderfer et al., 2018; Zupančič et al., 2022; Aguiar et al., 2026; Gilman & Baglietto, 2017). Such heat flux partitioning models remain central to wall boiling closures in multiphase computations or simulations, where the total wall heat flux is represented as a sum of mechanism-level contributions governed by empirical or semi-empirical closure laws (Liu & Dinh, 2018; Yeoh & Tu, 2009; Favre et al., 2026). High-fidelity simulations and experiments have further emphasized the importance of resolving microlayer, contact line, and dry spot dynamics, as well as conjugate heat transfer, for accurate boiling predictions (Bucci et al., 2016; Gerardi et al., 2010; Long et al., 2025). However, these approaches require detailed physical state variables, prescribed closure correlations, or expensive simulations. PAPA-Net takes a complementary data-driven route: it learns the mechanism partition directly from optical observations and sparse teacher masks, producing dry area, μ-layer, and single phase heat flux components without requiring masks or closure law parameters at inference.

Weak supervision, interpretable decomposition, and mechanism alignment. A related body of work studies how weak supervision and architectural constraints can make learned representations more interpretable or disentangled (Higgins et al., 2017; Locatello et al., 2019; 2020a). Object-centric and slot-based models similarly seek to decompose scenes into reusable parts or latent objects (Burgess et al., 2019; Greff et al., 2019; Locatello et al., 2020b), while concept-bottleneck and self-explaining models expose intermediate variables designed to align with human-interpretable concepts (Koh et al., 2020; Alvarez Melis & Jaakkola, 2018). However, these methods generally target semantic factors, objects, or concepts, rather than heat transfer mechanisms tied to the physical energy balance. Weakly supervised segmentation methods can learn masks from limited labels (Bearman et al., 2016; Zhu et al., 2023), but they do not by themselves enforce that each mask carries a physically meaningful heat contribution. PAPA-Net combines these ideas in a physics-aligned manner: sparse teacher masks assign mechanism identity, the mask-amplitude product enforces a physically interpretable heat flux partition, and optical-only inference preserves deployability without requiring phase annotations at test time.

3 METHOD

Intuitively, PAPA-Net asks each mechanism branch to answer two coupled questions: where is this mechanism active, and how much heat does it carry there? We formalize mechanism-resolved heat flux partitioning in §3.1, introduce the mask-amplitude factorization in §3.2, describe weak-teacher training in §3.3, and summarize optical-only inference in §3.4.

3.1 PROBLEM FORMULATION

We consider optical-only heat flux partitioning in boiling. Given an optical observation $x \in \mathbb{R}^{H\times W\times C}$, the goal is to predict both the total wall heat flux map $q'' \in \mathbb{R}^{H\times W}$ and a mechanism-resolved decomposition into $K = 3$ physical heat transfer components: dry area, μ-layer, and single phase. We write the predicted heat flux map as

$$\hat{q}'' = \sum_{k=1}^{K} \hat{q}''_k, \qquad k \in \{\mathrm{dry}, \mu\mathrm{-layer}, \mathrm{sp}\}. \tag{1}$$

The central challenge is that the total heat flux map alone does not identify the physical mechanism that produced each pixel. A model can fit $\hat{q}''$ while assigning heat to arbitrary or collapsed latent components. PAPA-Net addresses this by explicitly factorizing each component into a phase support mask and a nonnegative heat flux amplitude field.

3.2 PAPA-NET ARCHITECTURE

PAPA-Net uses a shared U-Net-style encoder-decoder backbone $F_\theta(x)$ followed by two per-mechanism prediction heads: a phase head that predicts spatial support masks and an amplitude

head that predicts nonnegative heat flux amplitude fields. The phase head predicts logits $\ell_k(x)$ for each mechanism and converts them into competing soft phase masks:

$$\hat{z}_k(x) = \frac{\exp(\ell_k(x)/\tau)}{\sum_{\ell=1}^{K} \exp(\ell_\ell(x)/\tau)}, \qquad \sum_{k=1}^{K} \hat{z}_k(x) = 1. \tag{2}$$

The temperature τ controls mask sharpness and is geometrically annealed during training; full schedule details are provided in Appendix A.3. The amplitude head predicts nonnegative per-mechanism heat flux amplitude fields:

$$\hat{a}_k(x) = \mathrm{softplus}(g_k(x)). \tag{3}$$

Each physical heat flux component is then formed by the mask-amplitude product

$$\hat{q}''_k(x) = \hat{z}_k(x)\,\hat{a}_k(x), \qquad \hat{q}''(x) = \sum_{k=1}^{K} \hat{z}_k(x)\hat{a}_k(x). \tag{4}$$

This factorization separates where a mechanism is active, $\hat{z}_k$, from how much heat it carries, $\hat{a}_k$. The softmax masks encourage phase competition, while the nonnegative amplitudes prevent unphysical signed cancellation between components.

3.3 Weak Teacher Physical Supervision

During training, a subset of frames may include teacher phase masks $m = \{m_k\}_{k=1}^{K}$. These masks are used only as supervision targets; they are never concatenated to the optical input of PAPA-Net. The local heat flux reconstruction loss is the Huber loss (Huber, 1992),

$$\mathcal{L}_{\mathrm{flux}} = \mathrm{Huber}_\delta(\hat{q}'', q''), \qquad \delta = 0.05. \tag{5}$$

To assign physical identity to the components, we supervise the predicted phase masks with a teacher-mask loss:

$$\mathcal{L}_{\mathrm{mask}} = \mathcal{L}_{\mathrm{Dice}}(\hat{z}, m) + \mathcal{L}_{\mathrm{CE}}(\hat{z}, m). \tag{6}$$

We also encourage the predicted heat contribution ratios to match teacher-derived contribution ratios. Let

$$\hat{c}_k = \frac{\sum_{i,j} \hat{q}''_k(i,j)}{\sum_{\ell=1}^{K}\sum_{i,j} \hat{q}''_\ell(i,j) + \epsilon}, \qquad c_k = \frac{\sum_{i,j} m_k(i,j) q''(i,j)}{\sum_{i,j} q''(i,j) + \epsilon}. \tag{7}$$

The contribution ratio loss is

$$\mathcal{L}_{\mathrm{contrib}} = \frac{1}{K}\sum_{k=1}^{K} \mathrm{Huber}_\delta(\hat{c}_k, c_k), \qquad \delta = 0.05. \tag{8}$$

The full training objective combines local heat-flux reconstruction, an auxiliary direct heat-flux prediction, weak-teacher supervision, contribution consistency, integrated heat-balance consistency, and regularization, with teacher-dependent terms evaluated only on samples with valid teacher masks:

$$\mathcal{L} = \lambda_{\mathrm{flux}}\mathcal{L}_{\mathrm{flux}} + \lambda_{\mathrm{direct}}\mathcal{L}_{\mathrm{direct}} + \lambda_{\mathrm{mask}}\mathcal{L}_{\mathrm{mask}} + \lambda_{\mathrm{contrib}}\mathcal{L}_{\mathrm{contrib}} + \lambda_{\mathrm{energy}}\mathcal{L}_{\mathrm{energy}} + \mathcal{L}_{\mathrm{reg}}. \tag{9}$$

Here, $\mathcal{L}_{\mathrm{direct}}$ is an auxiliary Huber heat-flux loss used only for training stabilization, while $\mathcal{L}_{\mathrm{reg}}$ collects the individually weighted priors and regularizers detailed in Appendix A.6. The global integrated heat-balance loss $\mathcal{L}_{\mathrm{energy}}$ penalizes mismatch in integrated heat flux, with $\epsilon = 10^{-8}$ used for numerical stability:

$$\mathcal{L}_{\mathrm{energy}} = \left| \frac{\sum_{i,j} \hat{q}''(i,j) - \sum_{i,j} q''(i,j)}{\sum_{i,j} q''(i,j) + \epsilon} \right|, \tag{10}$$

Algorithm 1 summarizes training, where $h_n \in \{0,1\}$ indicates a valid teacher mask and x_{n+1} is used only for temporal regularization.

Algorithm 1 PAPA-Net training and optical-only inference.

Require: training set $\mathcal{D}$; mechanisms $K = 3$; loss weights λ; temperature schedule τ_e
Ensure: Trained PAPA-Net parameters θ

for epoch $e = 1, \ldots, E$ **do**
  Set softmax temperature $\tau \leftarrow \tau_e$
  for minibatch $(x, q'', m, h, x_{\text{next}}) \sim \mathcal{D}$ **do**
    Encode optical input with shared backbone: $u \leftarrow F_\theta(x)$
    Predict heads: $(\ell_{1:K}, g_{1:K}, \hat{q}''_{\text{direct}}) \leftarrow H_\theta(u)$
    Compute phase masks and amplitudes: $\hat{z}_k = \text{softmax}_k(\ell/\tau)$, $\hat{a}_k = \text{softplus}(g_k)$
    Form predictions: $\hat{q}''_k = \hat{z}_k \hat{a}_k$, $\hat{q}'' = \sum_{k=1}^{K} \hat{q}''_k$
    Initialize $\mathcal{L} \leftarrow \lambda_{\text{flux}} \mathcal{L}_{\text{flux}}(\hat{q}'', q'') + \lambda_{\text{direct}} \mathcal{L}_{\text{direct}}(\hat{q}''_{\text{direct}}, q'') + \lambda_{\text{energy}} \mathcal{L}_{\text{energy}}(\hat{q}'', q'')$
    if there are labeled samples $\mathcal{B}_m = \{i : h_i = 1\}$ **then**
      Add $\lambda_{\text{mask}} \mathcal{L}_{\text{mask}}(\hat{z}, m) + \lambda_{\text{contrib}} \mathcal{L}_{\text{contrib}}(\hat{q}''_{1:K}, m, q'')$ on $\mathcal{B}_m$ to $\mathcal{L}$
    end if
    $\mathcal{L} \leftarrow \mathcal{L} + \mathcal{L}_{\text{reg}}(\hat{z}, \hat{q}'', x_{\text{next}})$
    Update θ by gradient descent on $\mathcal{L}$
  end for
end for

Optical only inference.

Require: Optical input x
Predict $\{\hat{z}_k, \hat{a}_k, \hat{q}''_k\}_{k=1}^{K}$ and $\hat{q}'' = \sum_{k=1}^{K} \hat{q}''_k$
return $\hat{q}''$ (total heat flux) and $\{\hat{q}''_k\}_{k=1}^{K}$ (dry, μ-layer, single phase)

3.4 Optical Only Inference and Relation to Baselines

At inference, PAPA-Net receives only the optical input x and outputs $\{\hat{z}_k, \hat{a}_k, \hat{q}''_k\}_{k=1}^{K}$ together with the total heat flux map $\hat{q}''$. No teacher masks or manual phase annotations are required. This distinguishes PAPA-Net from the Fully Guided Oracle, which uses ground-truth masks at inference and therefore serves only as an upper bound.

PAPA-Net also differs from cascade baselines. A cascade first predicts phase masks and then performs heat regression in a separate stage. In contrast, PAPA-Net learns masks and heat flux amplitudes jointly through Eq. (4), so the phase support and mechanism heat flux contributions are optimized end-to-end under the same heat-balance and contribution constraints. This joint factorization, rather than the use of masks as auxiliary inputs, is the core inductive bias tested in §4.3.

4 Experiments

This section evaluates PAPA-Net from the experimental protocol to physical diagnostics. §4.1 describes the boiling dataset, splits, baselines, and metrics; §4.2 reports IID and held-out sequence performance; §4.3 tests whether the mask-amplitude factorization aligns learned branches with physical mechanisms; §4.4 studies sparse teacher-mask supervision; and §4.5 evaluates physics-aware heat transfer metrics, held-out sequence generalization, and robustness.

4.1 Experimental Setup

Dataset and splits. Experiments use dT5-Boiling, a synchronized deionized-water flow boiling dataset acquired at 1.21 bar, mass flux $500\ \text{kg/m}^2\text{/s}$, and $5°\text{C}$ subcooling. The available applied heat flux conditions span $q'' = 0.49$–$2.17\ \text{MW/m}^2$, corresponding to $q''/q''_{\text{CHF}} = 0.22$–$0.99$, where the critical heat flux (CHF) is denoted q''_{CHF}. Each sample pairs an optical (e.g., background-subtracted phase detection frame) with an IR heat flux map and teacher phase masks for dry area, μ-layer, and single phase regions. We use sequences with $q'' \in \{0.49, 1.14, 1.53, 1.68, 1.94, 2.17\}\ \text{MW/m}^2$ as the IID sequence pool, with disjoint train, validation, and IID test frames. The $q'' = 1.82\ \text{MW/m}^2$ condition ($q''/q''_{\text{CHF}} = 0.83$, source S9) is reserved entirely for held-out sequence evaluation and is not used for training, validation, model selection, or calibration. Full acquisition conditions,

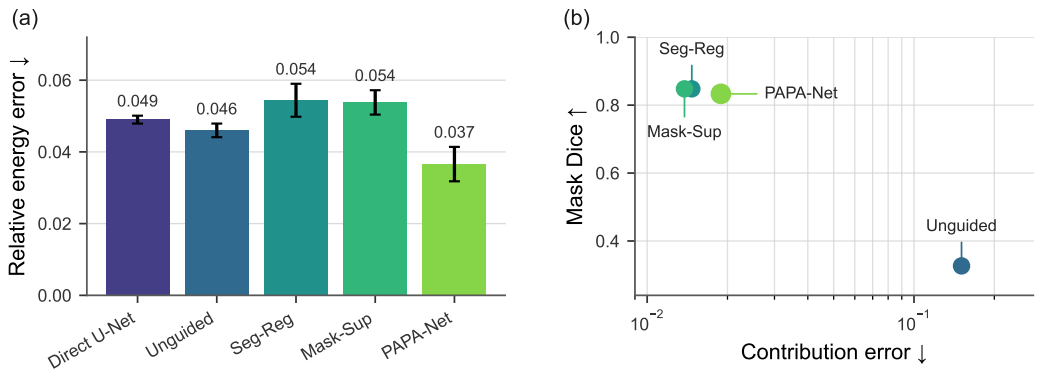


Figure 1: Multi-seed held-out performance of core optical-only methods on the held-out $q'' = 1.82\ \mathrm{MW/m^2}$ condition. Values are mean $\pm$ standard deviation over three seeds. (a) PAPA-Net achieves the lowest integrated heat error among optical-only methods. (b) Mechanism-quality Pareto plot comparing contribution error and mask Dice for methods with partition outputs. Direct U-Net is omitted from (b) because it has no mechanism decomposition.

sequence metadata, and the definition of dT5-Boiling are provided in Appendix A.1; the external dT10-Boiling dataset is described in §4.5 and Appendix B.7.

Teacher mask protocol. Teacher masks are used only as sparse training targets, never as inference inputs for PAPA-Net. At test time, PAPA-Net receives only the optical input and predicts the total heat flux map together with mechanism-resolved components. Fully Guided Oracle is the only method that receives ground-truth masks at inference, and is therefore treated as a mask-at-test upper bound rather than a deployable optical-only model. This protocol separates weak-teacher supervision from oracle access to physical masks.

Baselines. We choose baselines to isolate heat flux map regression, oracle partitioning, unguided decomposition, and practical optical-only partitioning. Direct U-Net is a U-Net-style optical baseline (Ronneberger et al., 2015) that predicts only q'' and tests whether accurate heat flux maps alone are sufficient. Fully Guided Oracle receives ground-truth phase masks at inference and measures the upper bound when physical regions are known at test time. Unguided uses the decomposition architecture without teacher-mask guidance, testing whether latent branches become physically meaningful on their own. Seg-Reg Cascade first trains a mask predictor from optical input and then uses the predicted masks for heat flux regression. Mask-Sup Cascade is a stronger supervised cascade variant with explicit teacher-mask supervision during training. Both cascades remain optical-only at inference because they use predicted masks, not ground-truth masks. PAPA-Net differs by learning masks and nonnegative heat amplitudes jointly through $\hat{q}''_k = \hat{z}_k \hat{a}_k$, rather than treating mask prediction and heat regression as separate stages. These baselines are intentionally strong because PAPA-Net should not only beat naive decomposition; it should justify its end-to-end mask-amplitude factorization against obvious supervised pipelines.

Metrics. For total heat-flux prediction, we report MAE, RMSE, RelE, and SSIM (Wang et al., 2004); for physical partitioning, we report Dice, CErr, Active, and Collapse. Direct U-Net has no partition metrics. These metrics distinguish local heat-flux accuracy from mechanism alignment; definitions and implementation details are provided in Appendix A.

4.2 Main Results: Optical Only Heat Partitioning

Figure 1 compares PAPA-Net with optical-only heat-flux prediction, unguided decomposition, and strong supervised cascade baselines on the held-out $q'' = 1.82\ \mathrm{MW/m^2}$ condition.

The multi-seed results in Figure 1 confirm the central tradeoff. PAPA-Net is not the strongest unconstrained local heat-map regressor, but it achieves the lowest held-out integrated heat error among the core optical-only methods while maintaining noncollapsed mechanism partitions. On held-out $q'' = 1.82\ \mathrm{MW/m^2}$, PAPA-Net obtains RelE 0.0366 ± 0.0048, compared with 0.0490 ± 0.0011

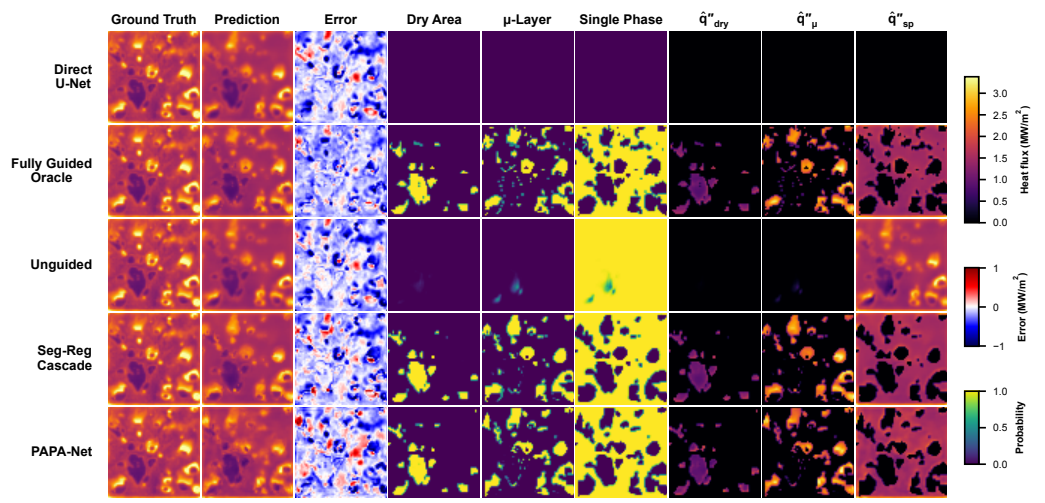


Figure 2: Qualitative mechanism diagnostics on the held-out $q'' = 1.82\,\mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$) condition. Direct U-Net predicts total heat flux but has no mechanism decomposition. Fully Guided Oracle uses ground-truth masks at inference and acts as an oracle upper bound. Unguided decomposition predicts plausible heat flux maps but collapses or misassigns physical branches. Seg-Reg Cascade is a strong supervised optical-only baseline. PAPA-Net remains optical-only at inference while producing dry area, μ-layer, and single phase masks and component-based heat flux maps.

for Direct U-Net, 0.0460 ± 0.0019 for Unguided, 0.0544 ± 0.0046 for Seg-Reg Cascade, and 0.0538 ± 0.0034 for Mask-Sup Cascade. Full multi-seed values are reported in Appendix Table 10, and bootstrap confidence intervals including the Fully Guided Oracle mask-at-test upper bound are reported in Appendix A.8.

Although the cascade baselines achieve slightly higher Dice or CErr on some splits, they solve a different two-stage problem: predict phase masks, then use them as auxiliary inputs for heat regression. This can improve segmentation-like scores, but it does not force each heat component to be a physically gated contribution. PAPA-Net instead learns a single coupled heat partition model in which each support mask directly gates a nonnegative mechanism amplitude, $\hat{q}''_k = \hat{z}_k \hat{a}_k$. Thus, the same components must jointly explain total heat balance, contribution ratios, and mechanism masks, yielding a deployable optical-only partition rather than segmentation-assisted regression. PAPA-Net is also lighter at inference, with runtime and parameter counts reported in Appendix C.

Figure 2 visualizes the mechanism-level behavior behind the quantitative results. Direct U-Net and unguided decomposition can produce plausible total heat flux maps, but they either lack physical components or collapse most heat into the single phase component. In contrast, PAPA-Net recovers spatially localized dry area and μ-layer contributions while preserving a coherent single phase background, closely matching the oracle structure without requiring masks at inference.

4.3 Why the Mask-Amplitude Factorization Aligns Physical Mechanisms

Figure 3 opens the PAPA-Net prediction into its two learned factors on the held-out $q'' = 1.82\,\mathrm{MW/m^2}$ condition. For each mechanism, the model predicts a spatial support mask $\hat{z}_k$ and a nonnegative amplitude field $\hat{a}_k$, whose product gives the component heat flux $\hat{q}''_k = \hat{z}_k \hat{a}_k$. We compare each predicted component against the teacher-derived reference partition $q''_{k,\mathrm{ref}} = m_k q''$. The reference component maps $q''_{k,\mathrm{ref}} = m_k q''$ are teacher-derived diagnostic proxies, not independently measured per-mechanism heat fluxes; they are used to evaluate whether predicted branches align with physically labeled phase regions. This diagnostic tests whether the learned branches are mechanism-aligned with dry area, μ-layer, and single phase regions rather than arbitrary latent channels. The corresponding IID diagnostic is provided in Appendix Figure 7.

Complementing the visual diagnostic in Figure 3, Figure 4 ablates the architectural and supervision choices that make PAPA-Net mechanism-aligned. The key observation is that strong heat flux map

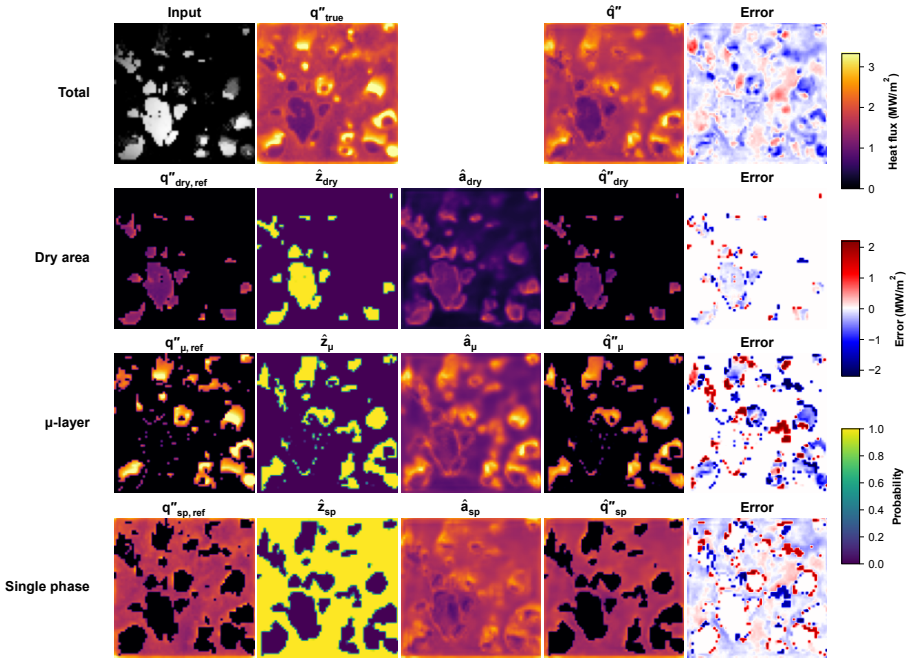


Figure 3: PAPA-Net factorization diagnostic on the held-out $q'' = 1.82\ \mathrm{MW/m^2}$ condition. At inference PAPA-Net receives only the optical input and predicts mechanism supports $\hat{z}_k$, nonnegative amplitudes $\hat{a}_k$, and components $\hat{q}''_k = \hat{z}_k \hat{a}_k$. Reference components are teacher-derived diagnostic proxies $q''_{k,\mathrm{ref}} = m_k q''$, not independently measured mechanism heat fluxes; error maps show $\hat{q}''_k - q''_{k,\mathrm{ref}}$.

accuracy does not imply physically meaningful partitioning. Independent component heat flux maps and no phase variants can fit total heat flux maps, but fail to recover mechanism identity, producing low Dice, large contribution error, inactive components, or full collapse. Similarly, removing teacher-mask guidance preserves competitive heat flux map error but destroys physical identity. These results show that the decomposition $\hat{q}''_k = \hat{z}_k \hat{a}_k$ is not merely a cosmetic latent decomposition: the phase masks provide physical support, the amplitudes model mechanism-specific heat intensity, and weak-teacher guidance assigns stable physical identity to the branches.

Appendix A.8 reports frame-bootstrap confidence intervals for these ablations. The conclusion is unchanged: Dice alone is insufficient, and removing factorization, phase structure, or teacher-mask guidance worsens contribution accuracy, suppresses active branches, or causes collapse.

4.4 Weak Teacher Label Efficiency

Sparse teacher masks are sufficient to align the physical branches. Without them, PAPA-Net preserves competitive heat flux error but collapses physically. On held-out $q'' = 1.82\ \mathrm{MW/m^2}$, moving from 0% to 1% teacher-mask supervision increases Dice from 0.2836 to 0.8132, reduces CErr from 0.2025 to 0.0302, and eliminates collapse (1.0000 to 0.0000). Dense supervision further improves CErr, but the key transition is from no teacher masks to sparse guidance. Full curves and bootstrap intervals are reported in Appendix Figure 8 and Table 13.

4.5 Physics-Aware Evaluation and Robustness

Beyond pixel-level error, PAPA-Net achieves the lowest integrated heat error and highest heat-trend correlation among optical-only methods on the held-out $q'' = 1.82\ \mathrm{MW/m^2}$ condition; full physics-aware metrics are in Appendix B.4.

We further evaluate zero-shot transfer to dT10-Boiling, collected in the same facility at a different time with a different heater. It matches dT5-Boiling in fluid, pressure, and mass flux but differs

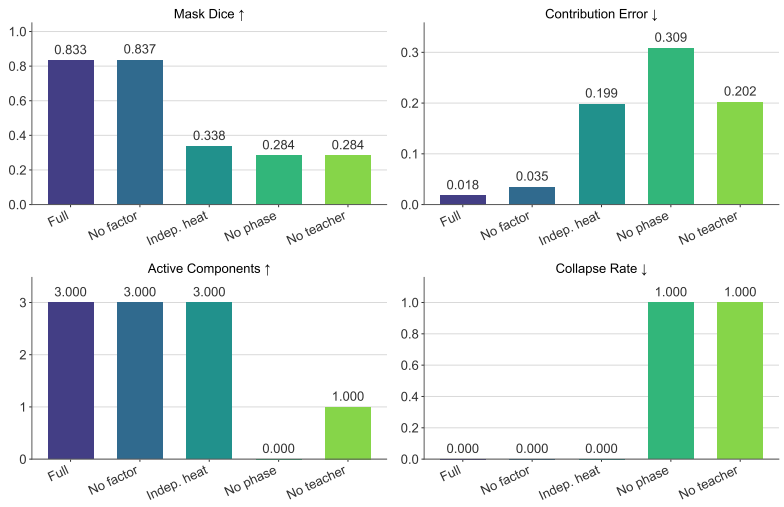


Figure 4: Mechanism alignment ablation on held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$). Competitive heat flux map prediction does not ensure meaningful mechanisms: removing phase structure or teacher guidance causes low Dice, high contribution error, inactive components, or collapse.

Table 1: Aggregate zero-shot dT10-Boiling results over 8 valid steps and 48,000 frames. Existing checkpoints are evaluated without retraining. Bold marks the best non-oracle, noncollapsed optical-only partitioning method for each metric; Fully Guided Oracle uses masks at inference and is shown only as an upper bound.

Method	RelE ↓	SSIM ↑	Dice ↑	CErr ↓	Active ↑	Collapse ↓
Direct U-Net	0.5358	0.3220	–	–	–	–
Fully Guided Oracle	0.4671	0.5054	1.0000	0.0197	2.669	0.223
Unguided	0.5104	0.3067	0.3546	0.1675	1.720	0.800
Seg-Reg Cascade	0.6434	0.2936	**0.6146**	0.0903	2.799	0.176
Mask-Sup Cascade	0.6426	0.2757	**0.6146**	0.0932	2.803	0.172
PAPA-Net	**0.5320**	**0.3285**	0.6111	**0.0783**	**2.807**	**0.166**

in subcooling and acquisition/post-processing conditions. Without retraining, Table 1 shows high absolute RelE, especially at low heat flux, while PAPA-Net has the strongest non-oracle mechanism-transfer profile among optical-only partitioning methods, with lower CErr, higher Active, and lower Collapse. Per-step curves and qualitative decompositions are in Appendix B.7.

Leave-one-condition-out evaluation shows that PAPA-Net preserves noncollapsed partitions across most internal heat-flux conditions; the low-energy $q'' = 0.49\ \mathrm{MW/m^2}$ regime remains the main failure case because dry-area events are rare. Robustness is stable under Gaussian noise and brightness/contrast shifts, with degradation mainly under blur and severe teacher-mask corruption (Appendices B.6 and B.5).

5 Conclusions

We introduce PAPA-Net for optical-only, mechanism-resolved wall heat flux partitioning in boiling. By factorizing heat flux components into physical support masks and nonnegative heat-flux amplitude fields, it produces dry area, μ-layer, and single phase contributions that sum to the total heat flux map. Across IID, held-out, and external zero-shot conditions, PAPA-Net produces mechanism-aligned, noncollapsed partitions and improves optical-only integrated heat balance. **Limitations and outlook.** PAPA-Net still uses teacher masks during training, and severe mask corruption or blurred optical inputs can degrade localization. Future work should test broader facilities, fluids, pressures, heat-flux regimes, and independent mechanism-level references.

AI USE STATEMENT

We used generative AI tools to polish sentence wording. All research ideas and scientific contributions are those of the authors, who take responsibility for the final content.

REFERENCES

Gustavo M. Aguiar, Matthew T. Hughes, and Matteo Bucci. On the nature of single-phase liquid heat transfer mechanisms in nucleate boiling: Evidence from high-resolution diagnostics. *International Journal of Multiphase Flow*, 198:105622, 2026. ISSN 0301-9322. doi: https://doi.org/10.1016/j.ijmultiphaseflow.2026.105622. URL https://www.sciencedirect.com/science/article/pii/S0301932226000236.

David Alvarez Melis and Tommi Jaakkola. Towards robust interpretability with self-explaining neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf.

Jan-Hendrik Bastek, WaiChing Sun, and Dennis Kochmann. Physics-informed diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=tpYeermigp.

Amy Bearman, Olga Russakovsky, Vittorio Ferrari, and Li Fei-Fei. What's the point: Semantic segmentation with point supervision. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (eds.), *Computer Vision – ECCV 2016*, pp. 549–565, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46478-7.

Matteo Bucci, Andrew Richenderfer, Guan-Yu Su, Thomas McKrell, and Jacopo Buongiorno. A mechanistic ir calibration technique for boiling heat transfer investigations. *International Journal of Multiphase Flow*, 83:115–127, 2016. ISSN 0301-9322. doi: https://doi.org/10.1016/j.ijmultiphaseflow.2016.03.007. URL https://www.sciencedirect.com/science/article/pii/S030193221530183X.

Christopher P. Burgess, Loic Matthey, Nicholas Watters, Rishabh Kabra, Irina Higgins, Matt Botvinick, and Alexander Lerchner. Monet: Unsupervised scene decomposition and representation, 2019. URL https://arxiv.org/abs/1901.11390.

Sanghyeon Chang, Youngjoon Suh, Chinmay Shingote, Cho-Ning Huang, Issam Mudawar, Chirag Kharangate, and Yoonjin Won. Bubblemask: Autonomous visualization of digital flow bubbles for predicting critical heat flux. *International Journal of Heat and Mass Transfer*, 217:124656, 2023. ISSN 0017-9310. doi: https://doi.org/10.1016/j.ijheatmasstransfer.2023.124656. URL https://www.sciencedirect.com/science/article/pii/S0017931023008013.

Victor H. Del Valle and D.B.R. Kenning. Subcooled flow boiling at high heat flux. *International Journal of Heat and Mass Transfer*, 28(10):1907–1920, 1985. ISSN 0017-9310. doi: https://doi.org/10.1016/0017-9310(85)90213-3. URL https://www.sciencedirect.com/science/article/pii/0017931085902133.

Luc Favre, Catherine Colin, Stéphane Pujet, and Stéphane Mimouni. On physical considerations regarding development and validation of heat flux partitioning models: Application to vertical boiling flows simulations. *International Journal of Heat and Mass Transfer*, 257:128235, 2026. ISSN 0017-9310. doi: https://doi.org/10.1016/j.ijheatmasstransfer.2025.128235. URL https://www.sciencedirect.com/science/article/pii/S0017931025015698.

Qianxi Fu, Youngjoon Suh, Xiaojing Zhang, Sanghyeon Chang, and Yoonjin Won. Bubble2heat: Optical to thermal inference in pool boiling using physics-encoded generative neural network. *AI Thermal Fluids*, 7:100043, 2026. ISSN 3050-5852. doi: https://doi.org/10.1016/j.aitf.2026.100043. URL https://www.sciencedirect.com/science/article/pii/S3050585226000170.

Aidan Furlong, Xingang Zhao, Robert K. Salko, and Xu Wu. Physics-based hybrid machine learning for critical heat flux prediction with uncertainty quantification. *Applied Thermal Engineering*, 279:127447, 2025. ISSN 1359-4311. doi: https://doi.org/10.1016/j.applthermaleng.2025.127447. URL https://www.sciencedirect.com/science/article/pii/S1359431125020393.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, November 2020. ISSN 2522-5839. doi: 10.1038/s42256-020-00257-z. URL https://doi.org/10.1038/s42256-020-00257-z.

Craig Gerardi, Jacopo Buongiorno, Lin wen Hu, and Thomas McKrell. Study of bubble growth in water pool boiling through synchronized, infrared thermometry and high-speed video. *International Journal of Heat and Mass Transfer*, 53(19):4185–4192, 2010. ISSN 0017-9310. doi: https://doi.org/10.1016/j.ijheatmasstransfer.2010.05.041. URL https://www.sciencedirect.com/science/article/pii/S0017931010002838.

Lindsey Gilman and Emilio Baglietto. A self-consistent, physics-based boiling heat transfer modeling framework for use in computational fluid dynamics. *International Journal of Multiphase Flow*, 95:35–53, 2017. ISSN 0301-9322. doi: https://doi.org/10.1016/j.ijmultiphaseflow.2017.04.018. URL https://www.sciencedirect.com/science/article/pii/S0301932215301774.

Klaus Greff, Raphaël Lopez Kaufman, Rishabh Kabra, Nick Watters, Christopher Burgess, Daniel Zoran, Loic Matthey, Matthew Botvinick, and Alexander Lerchner. Multi-object representation learning with iterative variational inference. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2424–2433. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/greff19a.html.

Katsiaryna Haitsiukevich, Onur Poyraz, Pekka Marttinen, and Alexander Ilin. Diffusion models as probabilistic neural operators for recovering unobserved states of dynamical systems. In *2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2024. doi: 10.1109/MLSP58920.2024.10734762.

Sheikh Md Shakeel Hassan, Arthur Feeney, Akash Dhruv, Jihoon Kim, Youngjoon Suh, Jaiyoung Ryu, Yoonjin Won, and Aparna Chandramowlishwaran. Bubbleml: A multiphase multiphysics dataset and benchmarks for machine learning. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 418–449. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/01726ae05d72ddba3ac784a5944fa1ef-Paper-Datasets_and_Benchmarks.pdf.

Sheikh Md Shakeel Hassan, Xianwei Zou, Akash Dhruv, and Aparna Chandramowlishwaran. Bubbleformer: Forecasting boiling with transformers. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38. Curran Associates, Inc., 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/18d2e6c115da385ff6a7ac53655ae95e-Paper-Datasets_and_Benchmarks_Track.pdf.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=Sy2fzU9gl.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. Diffusionpde: Generative pde-solving under partial observation. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 130291–130323. Curran Associates, Inc., 2024. doi:

10.52202/079017-4140. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/eb3878c1dcbfff9ee95d5d033e5f5942-Paper-Conference.pdf.

Peter J. Huber. Robust estimation of a location parameter. In Samuel Kotz and Norman L. Johnson (eds.), *Breakthroughs in Statistics: Methodology and Distribution*, pp. 492–518. Springer New York, New York, NY, 1992. ISBN 978-1-4612-4380-9. doi: 10.1007/978-1-4612-4380-9_35.

Christian Jacobsen, Yilin Zhuang, and Karthik Duraisamy. Cocogen: Physically consistent and conditioned score-based generative models for forward and inverse problems. *SIAM Journal on Scientific Computing*, 47(2):C399–C425, 2025. doi: 10.1137/24M1636071. URL https://doi.org/10.1137/24M1636071.

R. L. Judd and K. S. Hwang. A comprehensive model for nucleate pool boiling heat transfer including microlayer evaporation. *Journal of Heat Transfer*, 98(4):623–629, 11 1976. ISSN 0022-1481. doi: 10.1115/1.3450610. URL https://doi.org/10.1115/1.3450610.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, June 2021. ISSN 2522-5820. doi: 10.1038/s42254-021-00314-5. URL https://doi.org/10.1038/s42254-021-00314-5.

Siavash Khodakarami, Pouya Kabirzadeh, Chi Wang, Tarandeep Singh Thukral, and Nenad Miljkovic. Image-to-infrared mapping of vapor-to-liquid phase change dynamics using generative machine learning. *Cell Reports Physical Science*, 6(10):102845, 2025. ISSN 2666-3864. doi: https://doi.org/10.1016/j.xcrp.2025.102845. URL https://www.sciencedirect.com/science/article/pii/S2666386425004448.

Pang Wei Koh, Thao Nguyen, Yew Siang Tang, Stephen Mussmann, Emma Pierson, Been Kim, and Percy Liang. Concept bottleneck models. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 5338–5348. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/koh20a.html.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: learning maps between function spaces with applications to pdes. *J. Mach. Learn. Res.*, 24(1), January 2023. ISSN 1532-4435.

N. Kurul and Michael Z. Podowski. Multidimensional effects in forced convection subcooled boiling. In *Proceedings of the International Heat Transfer Conference 9*, volume 2, pp. 21–26, Jerusalem, Israel, 1990. Begell House. ISBN 978-0-89116-909-3. doi: 10.1615/IHTC9.40. URL https://ihtcdigitallibrary.com/conferences/6ec9fdc764f29109,3b6668e93c4d03e5,1968e4d71e4e45b1.html.

Zeyu Li, Wang Han, Yue Zhang, Qingfei Fu, Jingxuan Li, Lizi Qin, Ruoyu Dong, Hao Sun, Yue Deng, and Lijun Yang. Learning spatiotemporal dynamics with a pretrained generative model. *Nature Machine Intelligence*, 6(12):1566–1579, December 2024. ISSN 2522-5839. doi: 10.1038/s42256-024-00938-z. URL https://doi.org/10.1038/s42256-024-00938-z.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Yang Liu and Nam Dinh. Validation and uncertainty quantification for wall boiling closure relations in multiphase-cfd solver, 2018. URL https://arxiv.org/abs/1809.09184.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4114–4124. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/locatello19a.html.

Francesco Locatello, Ben Poole, Gunnar Raetsch, Bernhard Schölkopf, Olivier Bachem, and Michael Tschannen. Weakly-supervised disentanglement without compromises. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6348–6359. PMLR, 13–18 Jul 2020a. URL https://proceedings.mlr.press/v119/locatello20a.html.

Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 11525–11538. Curran Associates, Inc., 2020b. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/8511df98c02ab60aea1b2356c013bc0f-Paper.pdf.

Tian Long, Jieyun Pan, Edoardo Cipriano, Matteo Bucci, and Stéphane Zaleski. Direct numerical simulation of nucleate boiling with a resolved microlayer and conjugate heat transfer. *Journal of Fluid Mechanics*, 1020:A30, 2025. doi: 10.1017/jfm.2025.10542.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, March 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00302-5. URL https://doi.org/10.1038/s42256-021-00302-5.

Andrei Margeloiu, Matthew Ashman, Umang Bhatt, Yanzhi Chen, Mateja Jamnik, and Adrian Weller. Do concept bottleneck models learn as intended? *CoRR*, 2021. doi: 10.17863/CAM.80941. URL https://www.repository.cam.ac.uk/handle/1810/333521. ICLR 2021 Workshop on Responsible AI.

B. B. Mikic and W. M. Rohsenow. A new correlation of pool-boiling data including the effect of heating surface characteristics. *Journal of Heat Transfer*, 91(2):245–250, 05 1969. ISSN 0022-1481. doi: 10.1115/1.3580136. URL https://doi.org/10.1115/1.3580136.

M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

Andrew Richenderfer, Artyom Kossolapov, Jee Hyun Seong, Giacomo Saccone, Etienne Demarly, Ravikishore Kommajosyula, Emilio Baglietto, Jacopo Buongiorno, and Matteo Bucci. Investigation of subcooled flow boiling and chf using high-resolution diagnostics. *Experimental Thermal and Fluid Science*, 99:35–58, 2018. ISSN 0894-1777. doi: https://doi.org/10.1016/j.expthermflusci.2018.07.017. URL https://www.sciencedirect.com/science/article/pii/S089417771831255X.

W. M. Rohsenow. A method of correlating heat-transfer data for surface boiling of liquids. *Transactions of the American Society of Mechanical Engineers*, 74(6):969–975, 07 2022. ISSN 0097-6822. doi: 10.1115/1.4015984. URL https://doi.org/10.1115/1.4015984.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

Andrew Slavin Ross, Michael C. Hughes, and Finale Doshi-Velez. Right for the right reasons: Training differentiable models by constraining their explanations. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 2662–2670, 2017. doi: 10.24963/ijcai.2017/371. URL https://doi.org/10.24963/ijcai.2017/371.

Jee Hyun Seong, Madhumitha Ravichandran, Guanyu Su, Bren Phillips, and Matteo Bucci. Automated bubble analysis of high-speed subcooled flow boiling images using u-net transfer learning and global optical flow. *International Journal of Multiphase Flow*, 159:104336, 2023. ISSN 0301-9322. doi: https://doi.org/10.1016/j.ijmultiphaseflow.2022.104336. URL https://www.sciencedirect.com/science/article/pii/S0301932222002956.

Mehdi Shadkhah, Ronak Tali, Ali Rabeh, Cheng-Hau Yang, Ethan Herron, Abhisek Upadhyaya, Adarsh Krishnamurthy, Chinmay Hegde, Aditya Balu, and Baskar Ganapathysubramanian. MPFBench: A large scale dataset for sciML of multi-phase-flows: Droplet and bubble dynamics. *Journal of Data-centric Machine Learning Research*, 2025. ISSN XXXX-XXXX. URL https://openreview.net/forum?id=9gQnIFI4es.

Youngjoon Suh, Aparna Chandramowlishwaran, and Yoonjin Won. Recent progress of artificial intelligence for liquid-vapor phase change heat transfer. *npj Computational Materials*, 10(1):65, March 2024. ISSN 2057-3960. doi: 10.1038/s41524-024-01223-8. URL https://doi.org/10.1038/s41524-024-01223-8.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022. ISSN 0309-1708. doi: https://doi.org/10.1016/j.advwatres.2022.104180. URL https://www.sciencedirect.com/science/article/pii/S0309170822000562.

Guan Heng Yeoh and Jiyuan Tu. *Computational Techniques for Multiphase Flows: Basics and Applications*. Butterworth-Heinemann, 2009. ISBN 978-0-08-046733-7. doi: 10.1016/C2009-0-16604-7. URL https://www.sciencedirect.com/book/monograph/9780080467337/computational-techniques-for-multiphase-flows.

Kaiyin Zhu, Neal N. Xiong, and Mingming Lu. A survey of weakly-supervised semantic segmentation. In *2023 IEEE 9th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pp. 10–15, 2023. doi: 10.1109/BigDataSecurity-HPSC-IDS58521.2023.00013.

Matevž Zupančič, Peter Gregorčič, Mattia Bucci, Chi Wang, Gustavo Matana Aguiar, and Matteo Bucci. The wall heat flux partitioning during the pool boiling of water on thin metallic foils. *Applied Thermal Engineering*, 200:117638, 2022. ISSN 1359-4311. doi: https://doi.org/10.1016/j.applthermaleng.2021.117638. URL https://www.sciencedirect.com/science/article/pii/S1359431121010644.

APPENDIX

Appendix Table of Contents

A Additional Experimental Details 16
A.1 Dataset details 16
A.2 Data preprocessing and teacher mask handling 16
A.3 Training and model selection 17
A.4 Architecture and regularization details 18
A.5 Baseline details 19
A.6 Final PAPA-Net loss weights and tuning sweep 19
A.7 Metric definitions and physics-aware diagnostics 20
A.8 Statistical uncertainty estimates 23
B Extended Baselines and Ablations 24
B.1 Full architecture ablation 25
B.2 Full loss ablation 26
B.3 Teacher-mask label efficiency 27
B.4 Full physics-aware metrics 28
B.5 Robustness to optical and teacher-mask corruptions 29
B.6 Leave one condition out robustness and low-flux outlier audit 30
B.7 dT10-Boiling zero-shot transfer 31
C Runtime and Deployment Analysis 31
D Controlled Synthetic Partition Benchmark 32
D.1 Synthetic data generation 33
D.2 Splits and purpose 34

Appendix Tables

Table 2 Flow boiling sequence metadata 16
Table 3 Shared acquisition conditions 17
Table 4 Preprocessing and data loading settings 18
Table 5 Training hyperparameters 18
Table 6 Baseline roles and inference inputs 19
Table 7 Final PAPA-Net loss weights 19
Table 8 Weak teacher tuning sweep 20
Table 9 Metric groups and diagnostics 20
Table 10 Full multi-seed main results 23
Table 11 Main bootstrap uncertainty 24
Table 12 Mechanism-alignment uncertainty 25
Table 13 Teacher-mask fraction uncertainty 25
Table 14 Architecture ablation 26
Table 15 Full loss ablation 27
Table 16 Teacher-mask label efficiency 27
Table 17 Physics-aware metrics 28
Table 18 Leave-one-condition-out robustness 30
Table 19 Low-flux outlier audit 31
Table 20 dT10-Boiling evaluated heat-flux steps 31
Table 21 Runtime and deployment analysis 32
Table 22 Controlled synthetic benchmark results 34

Appendix Figures

Figure 5 Dataset overview across applied heat-flux conditions 17
Figure 6 IID qualitative diagnostics 24
Figure 7 IID factorization diagnostic 26
Figure 8 Teacher-mask label efficiency 28
Figure 9 Physics-aware evaluation 29
Figure 10 Robustness to optical and teacher-mask corruptions 29
Figure 11 Leave one condition out robustness 30
Figure 12 dT10-Boiling per-step zero-shot transfer 32
Figure 13 dT10-Boiling dataset overview 32
Figure 14 dT10-Boiling qualitative diagnostics 33

Figure 15 Synthetic OOD qualitative results 35

Notation and Abbreviations

Symbol	Meaning	Symbol	Meaning
x	optical input frame	n, N	frame index, number of frames
(i, j)	pixel index	$H, W, \lvert\Omega\rvert$	map size and pixel count
$q''_n, \hat{q}''_n$	true/predicted heat-map	q''_{CHF}	critical heat flux
q''/q''_{CHF}	normalized heat flux ratio	K, k, ℓ	mechanism and dummy indices
Ω	pixel domain	ϵ	numerical stability constant
$\hat{z}_{n,k}$	predicted support mask	$m_{n,k}$	teacher phase mask
$\hat{a}_k$	predicted heat amplitude field	$\hat{q}''_{n,k}$	predicted component heat flux
h	teacher mask availability	τ	softmax temperature
λ	loss weight	$\mathbf{1}[\cdot]$	indicator function
$\mu_{(\cdot)}, \sigma_{(\cdot)}$	SSIM mean/std-cov terms	c_1, c_2	SSIM stabilizers
$\hat{y}_n, y_n$	predicted/teacher hard labels	$D(a, b)$	pairwise Dice for channels a, b
S_K	permutations of K mechanisms	$\pi^\star$	best channel matching permutation
$\hat{c}_{n,k}$	predicted contribution ratio	$c_{n,k}$	teacher-derived contribution ratio
$\hat{s}_{n,k}$	predicted support area ratio	$A_{n,k}$	active mechanism indicator
$\tau_{\mathrm{area}}, \tau_{\mathrm{contrib}}$	active thresholds	$Q_n, \hat{Q}_n$	true/predicted integrated heat
$\overline{Q}, \overline{\hat{Q}}$	mean integrated heat values	$\Omega^{q''}_{n,10\%}$	top-10% heat pixels
$\Omega^{\nabla q''}_{n,10\%}$	top-10% gradient pixels	a_k^{teacher}	teacher area ratio
r_k	teacher event rate	MAE/RMSE	heat flux map errors
RelE/IntE/AbsE	integrated heat errors	SSIM	structural similarity index
Dice/CErr	mask/contribution errors	Active/Collapse	branch degeneracy metrics
$\mathrm{MAE}_{\mathrm{top}}, \mathrm{MAE}_\nabla$	top heat/gradient errors	VarErr, ρ_Q	temporal variation/trend correlation
OOD	held-out/generalization split		

A Additional Experimental Details

A.1 Dataset Details

dT5-Boiling consists of synchronized deionized-water flow boiling acquisitions at fixed pressure, mass flux, and 5°C subcooling. Phase detection videos are synchronized with IR heat flux and temperature fields, and binary phase region masks provide teacher labels for dry area, μ-layer, and liquid/single phase regions. The model uses the mask channels dry, μ-layer, and liquid/single phase for the three physical mechanisms. Figure 5 visualizes representative optical inputs, heat-flux maps, and teacher mechanism masks across the applied heat-flux conditions. The name dT5-Boiling indicates the 5°C subcooling condition and distinguishes this primary dataset from the external dT10-Boiling transfer dataset.

Table 2: Flow boiling sequence metadata. q'' denotes applied heat flux and q''/q''_{CHF} is normalized by the measured critical heat flux $q''_{\mathrm{CHF}} = 2.2\ \mathrm{MW/m^2}$.

Source ID	Applied heat flux q''	CHF ratio q''/q''_{CHF}	Notes
S4	$0.493\ \mathrm{MW/m^2}$	0.224	Low flux outlier; rare dry area events
S6	$1.138\ \mathrm{MW/m^2}$	0.517	22 deleted frames recorded
S7	$1.529\ \mathrm{MW/m^2}$	0.695	Stable mid/high condition
S8	$1.684\ \mathrm{MW/m^2}$	0.765	11 deleted frames recorded
S9	$1.823\ \mathrm{MW/m^2}$	0.828	Held-out sequence
S11	$1.941\ \mathrm{MW/m^2}$	0.882	High flux condition
S12	$2.172\ \mathrm{MW/m^2}$	0.987	Near CHF condition

A.2 Data preprocessing and teacher mask handling

Each sample consists of an optical input x, heat flux target q'', optional teacher mask m, and a neighboring optical frame x_{next} used for temporal regularization. Teacher masks are used only as supervision; they are never concatenated to the input of PAPA-Net. Unless otherwise stated, inputs are resized to 64×64, the heat channel is normalized by 5×10^6, and teacher masks are binarized with threshold 0.5. We use three teacher mask channels corresponding to dry area, μ-layer, and single phase. For the real boiling experiments, the default mask channel order is dry/μ/single phase.

For both datasets, the water-side heat flux is obtained by subtracting the sapphire heat-flux contribution from the applied heat flux. For dT5-Boiling, this processing had already been applied in the

Table 3: Shared acquisition conditions.

Quantity	Value
Fluid	Deionized water
Configuration	Flow boiling
Pressure	1.21 bar
Mass flux	500 $\mathrm{kg/m^2/s}$
Bulk / saturation temperature	100.2°C / 105.2°C
Subcooling	5°C
Critical heat flux q''_{CHF}	2.2 $\mathrm{MW/m^2}$
Heater area	10^{-4} $\mathrm{m^2}$
Phase detection video	10000 FPS, 800 × 800 pixels, 12.2397 μm/px
IR camera	2500 FPS, 91 × 91 pixels, 107.5269 μm/px
Mask classes	dry, μ-layer, liquid/single phase, footprint
Used mask channels	dry, μ-layer, liquid/single phase

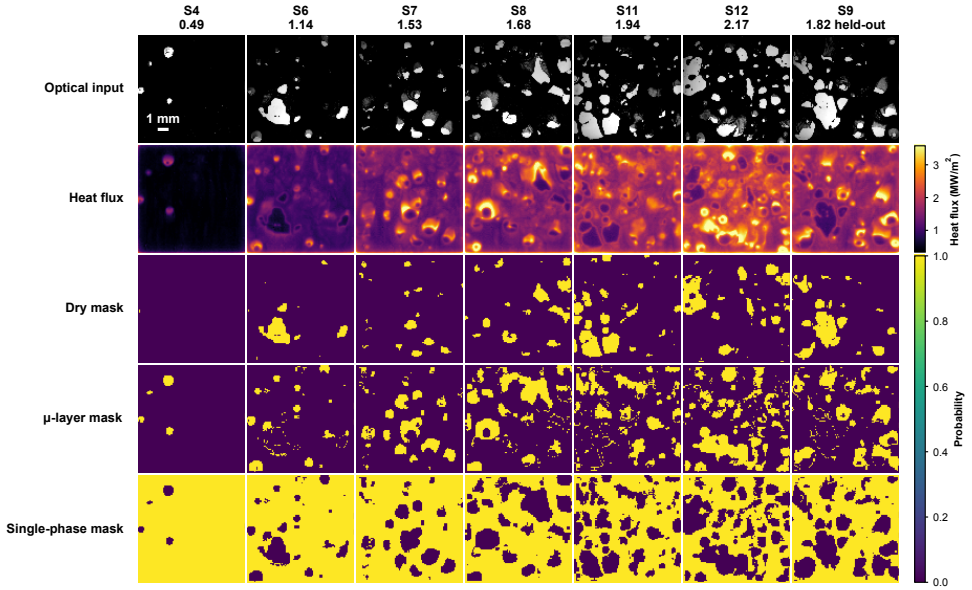


Figure 5: Dataset overview across applied heat-flux conditions. Each column shows one sequence; S9 is the held-out condition used for sequence-level generalization. Rows show the optical phase-detection input, corresponding IR heat-flux map, and teacher masks for dry area, μ-layer, and single phase. Teacher masks are used only during training and evaluation diagnostics; PAPA-Net inference uses only the optical input. The scale bar in the optical row denotes 1 mm.

provided IR files used by our pipeline. For dT10-Boiling, the supplied files required this subtraction during our data conversion.

A.3 Training and Model Selection

All neural models are trained with AdamW (Loshchilov & Hutter, 2019) using a cosine learning rate schedule. PAPA-Net uses geometric temperature annealing for the phase-mask softmax. At epoch $e \in \{1, \ldots, E\}$, the temperature is

$$\tau_e = \tau_{\mathrm{start}} \left(\frac{\tau_{\mathrm{end}}}{\tau_{\mathrm{start}}} \right)^{\frac{e-1}{E-1}}, \tag{11}$$

with $\tau_{\mathrm{start}} = 1.0$, $\tau_{\mathrm{end}} = 0.15$, and $E = 150$. Thus, the temperature decreases uniformly in log-space over training, progressively sharpening the competing mechanism masks. Unless otherwise stated, models are trained for 150 epochs with learning rate 2×10^{-4}. The final PAPA-Net runs use

Table 4: Default preprocessing and data loading settings.

Setting	Value
Input resolution	64×64 pixels
Heat channel	0
Mask channels	dry, μ-layer, single phase
Mask threshold	0.5
Input scale	0.6
Heat flux scale	5×10^6 W/m^2
Frame skip	1000
Frames per sequence	2000 unless otherwise noted
IID sequences	S4, S6, S7, S8, S11, S12
Held-out condition	$q'' = 1.82$ MW/m^2 ($q''/q''_{\mathrm{CHF}} = 0.83$, source S9)

batch size 192, while some baseline and sweep runs use the batch sizes listed in their configuration files. The best checkpoint is selected by validation MAE for heat regression models. The final PAPA-Net model used in the main paper is the validation-selected weak-teacher configuration from the tuning sweep in Appendix A.6. All Huber losses use $\delta = 0.05$ in normalized heat-flux units; since heat maps are scaled by 5×10^6, this corresponds to approximately 0.25 MW/m^2, giving quadratic penalties for typical residuals while reducing sensitivity to rare high-flux hotspot errors.

Hardware and execution. All experiments were run on a local workstation with an Intel Core i7-14700K CPU, 64 GiB RAM, and two NVIDIA GeForce RTX 4090 GPUs with 24 GB memory each. Each reported PAPA-Net training run used a single RTX 4090 in FP32 for 150 epochs; the main PAPA-Net runs used batch size 192 with 8 data-loading workers. Independent sweeps were parallelized across the two GPUs by assigning separate runs to separate devices. Inference and evaluation were performed on a single CUDA device selected automatically.

Table 5: Default training hyperparameters.

Hyperparameter	Value
Optimizer	AdamW
Learning rate	2×10^{-4}
Weight decay	1×10^{-4}
Batch size	192 final PAPA-Net; otherwise config-specific
Epochs	150
Validation fraction	0.15
Base channels	32
Number of phases	3
Softmax temperature start	1.0
Softmax temperature end	0.15
Softmax temperature schedule	Geometric decay over epochs
Random seed	42
Checkpoint selection	Best validation MAE

A.4 Architecture and Regularization Details

PAPA-Net uses a shared U-Net-style encoder-decoder backbone followed by two per-mechanism heads. The phase head predicts logits $\ell_k(x)$ for each physical branch and applies a temperature-scaled softmax to obtain spatially competing phase masks $\hat{z}_k$. The temperature is geometrically annealed from 1.0 to 0.15 using the schedule defined in Appendix A.3. The amplitude head predicts raw heat intensity maps $g_k(x)$ and applies a softplus nonlinearity to enforce nonnegative mechanism amplitudes $\hat{a}_k = \mathrm{softplus}(g_k)$. The component heat flux maps are then computed as $\hat{q}''_k = \hat{z}_k \hat{a}_k$ and summed to obtain the total prediction $\hat{q}'' = \sum_k \hat{q}''_k$.

The regularization term $\mathcal{L}_{\mathrm{reg}}$ contains four lightweight stabilizers. Entropy/sharpening discourages overly diffuse phase masks. Total variation penalties encourage spatial smoothness in both masks and component heat flux maps. Branch decorrelation discourages different mechanisms from learn-

ing duplicate spatial supports. Temporal consistency encourages masks for adjacent frames to remain stable when the optical input changes smoothly. These terms are not intended to replace heat flux supervision; they prevent degenerate decompositions while the heat flux and weak-teacher losses assign physical meaning to the branches.

For models with an auxiliary direct head, the direct prediction is used only as a training stabilizer through $\mathcal{L}_{\text{direct}}$; the mechanism-resolved prediction remains the factorized output $\hat{q}'' = \sum_k \hat{z}_k \hat{a}_k$.

A.5 Baseline Details

We summarize the baseline roles used in the main experiments. Direct U-Net is a q'' only optical baseline and does not produce mechanism partitions. Fully Guided Oracle concatenates or otherwise uses ground-truth phase masks at inference, so it is not a deployable optical-only method; it is included only as an upper bound. Unguided uses the same general decomposition family but removes teacher-mask guidance and physical regularization, testing whether branches self organize without supervision. Seg-Reg Cascade first trains a mask predictor from optical input and then uses the predicted masks for heat flux regression. Mask-Sup Cascade is a stronger supervised cascade variant with explicit teacher-mask supervision during training. Both cascades remain optical-only at inference because they use predicted masks, not ground-truth masks. PAPA-Net differs by learning masks and nonnegative heat amplitudes jointly through $\hat{q}''_k = \hat{z}_k \hat{a}_k$, rather than treating mask prediction and heat regression as separate stages.

Table 6: Baseline roles and inference inputs.

Method	Inference input	Partition output	Purpose
Direct U-Net	optical-only	none	tests heat flux regression without mechanisms
Fully Guided Oracle	optical + ground-truth masks	mask-guided components	upper bound with privileged mask access
Unguided	optical-only	latent components	tests whether branches self organize without teacher masks
Seg-Reg Cascade	optical-only	predicted mask-guided components	two-stage supervised mask then heat regression baseline
Mask-Sup Cascade	optical-only	predicted mask-guided components	stronger supervised cascade with explicit mask supervision
PAPA-Net	optical-only	$\hat{q}''_k = \hat{z}_k \hat{a}_k$	joint mechanism support and heat amplitude factorization

A.6 Final PAPA-Net Loss Weights

The final PAPA-Net model uses the tuned weak-teacher setting. The tuning increases the heat reconstruction and direct head weights relative to the base weak-teacher model, improving heat flux map accuracy while preserving noncollapsed mechanism partitions.

Table 7: Final PAPA-Net loss weights used for the main results.

Loss term	Weight
Local heat-flux reconstruction $\mathcal{L}_{\text{flux}}$	2.00
Auxiliary direct heat-flux reconstruction $\mathcal{L}_{\text{direct}}$	0.20
Integrated heat-balance consistency $\mathcal{L}_{\text{energy}}$	0.10
Teacher-mask supervision $\mathcal{L}_{\text{mask}}$	0.30
Teacher contribution supervision $\mathcal{L}_{\text{contrib teacher}}$	0.20
Contribution prior $\mathcal{L}_{\text{contribprior}}$	0.10
Area prior $\mathcal{L}_{\text{areaprior}}$	0.03
Entropy / sharpening regularizer	0.02
Mask decorrelation regularizer	0.03
Mask total variation	0.002
Component heat total variation	0.001
Temporal consistency	0.01
Edge alignment	0.00

Tuning sweep. Table 8 summarizes the weak-teacher tuning sweep used to select the final PAPA-Net model. Increasing the heat flux and direct head weights improves heat flux map accuracy and integrated heat consistency while preserving high Dice and zero collapse on held-out $q'' = 1.82\ \text{MW/m}^2$ ($q''/q''_{\text{CHF}} = 0.83$). We therefore use the setting with $\lambda_{\text{flux}} = 2.0$ and $\lambda_{\text{direct}} = 0.20$ as the final PAPA-Net model.

Table 8: PAPA-Net weak-teacher tuning sweep on held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$). The selected final model improves heat flux map accuracy and integrated heat balance while preserving noncollapsed physical partitions.

Model	MAE ↓	RMSE ↓	RelE ↓	Dice ↑	CErr ↓	Collapse ↓
Base weak-teacher	0.0448	0.0637	0.0356	0.8370	0.0185	0.0000
Intermediate variant	0.0462	0.0657	0.0376	0.8323	0.0221	0.0000
Mask weight 0.50	0.0480	0.0682	0.0354	0.8281	0.0198	0.0000
Mask weight 0.75	0.0493	0.0698	0.0394	0.8329	0.0220	0.0000
Mask weight 1.00	0.0504	0.0715	0.0423	0.8324	0.0200	0.0000
Flux/direct weights $(2.0, 0.10)$	0.0438	0.0623	0.0364	0.8329	0.0188	0.0000
Final tuned model ($\lambda_{\mathrm{flux}} = 2.0, \lambda_{\mathrm{direct}} = 0.20$)	**0.0422**	**0.0600**	**0.0336**	0.8335	0.0180	0.0000
Contribution weight 0.30	0.0464	0.0659	0.0424	0.8335	0.0195	0.0000
Contribution weight 0.50	0.0459	0.0652	0.0385	0.8292	0.0183	0.0000
Two stage variant	0.0485	0.0685	0.0410	0.8277	0.0213	0.0000

A.7 Metric Definitions

The metrics are chosen to separate three failure modes. MAE, RMSE, SSIM, and RelE evaluate total heat flux prediction. Dice, CErr, Active, and Collapse evaluate whether the predicted branches correspond to physical mechanisms rather than arbitrary latent components. The physics-aware metrics evaluate heat transfer validity beyond pixel error, including global heat balance, hotspot accuracy, interface gradient accuracy, temporal contribution stability, and agreement with the true integrated heat trend.

Table 9: Metric groups and what they test.

Metric group	Metrics	What it tests
Heat flux map accuracy	MAE, RMSE, SSIM	local agreement between $\hat{q}''$ and q''
Integrated heat balance	RelE, IntE	global conservation of total wall heat flux
Mechanism identity	Dice, CErr	whether branches match dry area, μ-layer, and single phase
Branch degeneracy	Active, Collapse	whether mechanisms remain active or collapse
Physics diagnostics	Top heat MAE, gradient MAE, VarErr, trend correlation	hotspot accuracy, interface accuracy, temporal stability, heat-trend agreement

We evaluate both total heat flux accuracy and mechanism-level physical alignment. Let $q''_n \in \mathbb{R}^{H\times W}$ and $\hat{q}''_n \in \mathbb{R}^{H\times W}$ denote the ground truth and predicted heat flux maps for frame n, with pixel domain Ω and $|\Omega| = HW$. We use $K = 3$ mechanisms: dry area, μ-layer, and single phase. Predicted mechanism masks and component heat flux maps are denoted by $\hat{z}_{n,k}$ and $\hat{q}''_{n,k}$, while teacher masks are denoted by $m_{n,k}$. A small constant ϵ is used for numerical stability.

Heat flux map accuracy. Mean absolute error and root mean squared error are defined as

$$\mathrm{MAE} = \frac{1}{N|\Omega|}\sum_{n=1}^{N}\sum_{(i,j)\in\Omega}\left|\hat{q}''_n(i,j) - q''_n(i,j)\right|. \tag{12}$$

$$\mathrm{RMSE} = \left[\frac{1}{N|\Omega|}\sum_{n=1}^{N}\sum_{(i,j)\in\Omega}\left(\hat{q}''_n(i,j) - q''_n(i,j)\right)^2\right]^{1/2}. \tag{13}$$

Relative integrated heat error measures global heat balance consistency:

$$\mathrm{RelE} = \frac{1}{N}\sum_{n=1}^{N}\frac{\left|\sum_{(i,j)\in\Omega}\hat{q}''_n(i,j) - \sum_{(i,j)\in\Omega}q''_n(i,j)\right|}{\left|\sum_{(i,j)\in\Omega}q''_n(i,j)\right| + \epsilon}. \tag{14}$$

For structural similarity, we compute global frame wise SSIM and average over the evaluation set:

$$\mathrm{SSIM} = \frac{1}{N}\sum_{n=1}^{N}\frac{(2\mu_{\hat{q}''_n}\mu_{q''_n} + c_1)(2\sigma_{\hat{q}''_n q''_n} + c_2)}{(\mu^2_{\hat{q}''_n} + \mu^2_{q''_n} + c_1)(\sigma^2_{\hat{q}''_n} + \sigma^2_{q''_n} + c_2)}. \tag{15}$$

Mechanism mask accuracy. Because mechanism channels can permute in unsupervised or weakly supervised decompositions, mask scores are computed after best channel matching. Let

$$\hat{y}_n(i,j) = \arg\max_k \hat{z}_{n,k}(i,j), \qquad y_n(i,j) = \arg\max_k m_{n,k}(i,j) \tag{16}$$

be the predicted and teacher phase labels. For predicted channel a and teacher channel b, define

$$D(a,b) = \frac{2\sum_{n=1}^{N}\sum_{(i,j)\in\Omega}\mathbf{1}[\hat{y}_n(i,j)=a]\mathbf{1}[y_n(i,j)=b]+\epsilon}{\sum_{n=1}^{N}\sum_{(i,j)\in\Omega}\mathbf{1}[\hat{y}_n(i,j)=a]+\sum_{n=1}^{N}\sum_{(i,j)\in\Omega}\mathbf{1}[y_n(i,j)=b]+\epsilon}. \tag{17}$$

The best permutation of predicted channels is

$$\pi^\star = \arg\max_{\pi\in S_K}\frac{1}{K}\sum_{k=1}^{K} D(\pi(k),k), \tag{18}$$

where S_K is the set of all permutations of K mechanisms. The reported Dice is

$$\text{Dice} = \frac{1}{K}\sum_{k=1}^{K} D(\pi^\star(k),k). \tag{19}$$

Mechanism contribution accuracy. Predicted heat contribution ratios are computed from the predicted component heat flux maps:

$$\hat{c}_{n,k} = \frac{\sum_{(i,j)\in\Omega}\hat{q}''_{n,k}(i,j)}{\sum_{\ell=1}^{K}\sum_{(i,j)\in\Omega}\hat{q}''_{n,\ell}(i,j)+\epsilon}. \tag{20}$$

Teacher derived contribution ratios are computed by assigning the true heat flux map to teacher phase regions:

$$c_{n,k} = \frac{\sum_{(i,j)\in\Omega} m_{n,k}(i,j)q''_n(i,j)}{\sum_{(i,j)\in\Omega} q''_n(i,j)+\epsilon}. \tag{21}$$

To remain consistent with the Dice evaluation, predicted contributions are matched using the best permutation $\pi^\star$. Contribution error is then

$$\text{CErr} = \frac{1}{NK}\sum_{n=1}^{N}\sum_{k=1}^{K}\left|\hat{c}_{n,\pi^\star(k)} - c_{n,k}\right|. \tag{22}$$

Collapse and active mechanisms. We define the predicted area ratio of mechanism k as

$$\hat{s}_{n,k} = \frac{1}{|\Omega|}\sum_{(i,j)\in\Omega}\hat{z}_{n,k}(i,j). \tag{23}$$

A mechanism is considered active if it occupies nontrivial area and carries nontrivial heat contribution:

$$A_{n,k} = \mathbf{1}\left[\hat{s}_{n,k} > \tau_{\text{area}}\right]\mathbf{1}\left[\hat{c}_{n,k} > \tau_{\text{contrib}}\right], \quad \tau_{\text{area}} = \tau_{\text{contrib}} = 0.01. \tag{24}$$

The reported number of active components is

$$\text{Active} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} A_{n,k}. \tag{25}$$

Collapse rate is the fraction of frames for which fewer than all K mechanisms are active:

$$\text{Collapse} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{1}\left[\sum_{k=1}^{K} A_{n,k} < K\right]. \tag{26}$$

Physics-aware metrics. For physics-aware evaluation, integrated heat error is equivalent to the relative error in total predicted heat:

$$\mathrm{IntE} = \frac{1}{N}\sum_{n=1}^{N}\frac{\left|Q_n - \hat{Q}_n\right|}{|Q_n| + \epsilon}, \quad Q_n = \sum_{(i,j)\in\Omega} q_n''(i,j), \quad \hat{Q}_n = \sum_{(i,j)\in\Omega} \hat{q}_n''(i,j). \tag{27}$$

The absolute integrated heat error used in the low-flux outlier audit is

$$\mathrm{AbsE} = \frac{1}{N}\sum_{n=1}^{N}\left|Q_n - \hat{Q}_n\right|. \tag{28}$$

Top-10% heat region MAE evaluates error on the highest heat pixels:

$$\mathrm{MAE}_{\mathrm{top}} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{|\Omega_{n,10\%}^{q''}|}\sum_{(i,j)\in\Omega_{n,10\%}^{q''}}\left|\hat{q}_n''(i,j) - q_n''(i,j)\right|, \tag{29}$$

where $\Omega_{n,10\%}^{q''}$ denotes the top 10% of pixels by ground truth heat flux in frame n.

High gradient region MAE evaluates error around sharp thermal transitions:

$$\mathrm{MAE}_{\nabla} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{|\Omega_{n,10\%}^{\nabla q''}|}\sum_{(i,j)\in\Omega_{n,10\%}^{\nabla q''}}\left|\hat{q}_n''(i,j) - q_n''(i,j)\right|, \tag{30}$$

where $\Omega_{n,10\%}^{\nabla q''}$ denotes the top 10% of pixels by the finite difference gradient magnitude of q_n''.

Temporal contribution stability compares frame to frame changes in mechanism contribution magnitude:

$$\mathrm{VarErr} = \frac{1}{(N-1)K}\sum_{n=2}^{N}\sum_{k=1}^{K}\left|\left|\hat{c}_{n,k} - \hat{c}_{n-1,k}\right| - \left|c_{n,k} - c_{n-1,k}\right|\right|. \tag{31}$$

Heat trend correlation measures temporal agreement between predicted and true integrated heat:

$$\rho_Q = \frac{\sum_{n=1}^{N}(Q_n - \overline{Q})(\hat{Q}_n - \overline{\hat{Q}})}{\sqrt{\sum_{n=1}^{N}(Q_n - \overline{Q})^2}\sqrt{\sum_{n=1}^{N}(\hat{Q}_n - \overline{\hat{Q}})^2} + \epsilon}. \tag{32}$$

For the low-flux outlier audit, we additionally report the mean true integrated heat

$$\overline{Q} = \frac{1}{N}\sum_{n=1}^{N} Q_n,$$

teacher area ratio

$$a_k^{\mathrm{teacher}} = \frac{1}{N|\Omega|}\sum_{n=1}^{N}\sum_{(i,j)\in\Omega} m_{n,k}(i,j),$$

and event rate

$$r_k = \frac{1}{N}\sum_{n=1}^{N}\mathbf{1}\left[\sum_{(i,j)\in\Omega} m_{n,k}(i,j) > 0\right].$$

These quantities diagnose whether a held-out sequence contains enough dry area, μ-layer, and single phase activity to excite all physical branches.

Table 10: Full multi-seed main results for core optical-only methods. Values are mean ± standard deviation over three seeds. This table provides the numeric values summarized in Figure 1.

Split	Method	MAE ↓	RMSE ↓	RelE ↓	SSIM ↑
IID	Direct U-Net	0.0283 ± 0.0006	0.0441 ± 0.0008	0.0110 ± 0.0007	0.8611 ± 0.0055
IID	Unguided	0.0286 ± 0.0005	0.0445 ± 0.0008	0.0112 ± 0.0002	0.8573 ± 0.0060
IID	Seg-Reg Cascade	0.0275 ± 0.0009	0.0428 ± 0.0020	0.0098 ± 0.0000	0.8725 ± 0.0123
IID	Mask-Sup Cascade	0.0268 ± 0.0016	0.0419 ± 0.0032	0.0105 ± 0.0029	0.8812 ± 0.0185
IID	PAPA-Net	0.0326 ± 0.0004	0.0508 ± 0.0007	0.0180 ± 0.0020	0.8237 ± 0.0041
Held-out	Direct U-Net	0.0377 ± 0.0005	0.0537 ± 0.0006	0.0490 ± 0.0011	0.8387 ± 0.0041
Held-out	Unguided	0.0379 ± 0.0006	0.0537 ± 0.0007	0.0460 ± 0.0019	0.8385 ± 0.0032
Held-out	Seg-Reg Cascade	0.0403 ± 0.0012	0.0569 ± 0.0011	0.0544 ± 0.0046	0.8259 ± 0.0019
Held-out	Mask-Sup Cascade	0.0395 ± 0.0009	0.0561 ± 0.0006	0.0538 ± 0.0034	0.8343 ± 0.0009
Held-out	PAPA-Net	0.0421 ± 0.0003	0.0598 ± 0.0004	0.0366 ± 0.0048	0.7948 ± 0.0067

Split	Method	Dice ↑	CErr ↓	Active ↑	Collapse ↓
IID	Direct U-Net	–	–	–	–
IID	Unguided	0.2971 ± 0.0207	0.2164 ± 0.0319	1.9330 ± 0.3505	0.7710 ± 0.0947
IID	Seg-Reg Cascade	0.8841 ± 0.0204	0.0090 ± 0.0032	2.7860 ± 0.0051	0.1752 ± 0.0010
IID	Mask-Sup Cascade	0.8841 ± 0.0204	0.0082 ± 0.0032	2.7850 ± 0.0073	0.1765 ± 0.0033
IID	PAPA-Net	0.8500 ± 0.0014	0.0173 ± 0.0002	2.7930 ± 0.0053	0.1715 ± 0.0040
Held-out	Direct U-Net	–	–	–	–
Held-out	Unguided	0.3267 ± 0.0330	0.1507 ± 0.0773	1.6270 ± 0.5458	1.0000 ± 0.0000
Held-out	Seg-Reg Cascade	0.8480 ± 0.0015	0.0147 ± 0.0013	3.0000 ± 0.0000	0.0000 ± 0.0000
Held-out	Mask-Sup Cascade	0.8480 ± 0.0015	0.0138 ± 0.0011	3.0000 ± 0.0000	0.0000 ± 0.0000
Held-out	PAPA-Net	0.8332 ± 0.0013	0.0189 ± 0.0010	3.0000 ± 0.0000	0.0000 ± 0.0000

A.8 Statistical uncertainty estimates

We report complementary uncertainty estimates for the main results and key mechanism-alignment studies. For the main comparison, we reran the core optical-only methods over three seeds and report mean ± standard deviation in Table 10. For completeness, we also report frame-bootstrap 95% confidence intervals over evaluation frames for the full main comparison, including the Fully Guided Oracle. Bootstrap intervals quantify evaluation-frame uncertainty for a fixed checkpoint and complement the multi-seed training-level uncertainty estimate. We use the same IID and held-out $q'' = 1.82\,\mathrm{MW/m^2}$ splits as in the main experiments.

Table 11: Frame-bootstrap 95% confidence intervals for the full main IID and held-out results. Entries are point estimate [lower, upper]. Fully Guided Oracle uses masks at inference and is included only as a mask-at-test upper bound.

Split	Method	MAE ↓	RMSE ↓	RelE ↓	SSIM ↑
IID	Direct U-Net	0.0278 [0.0273, 0.0283]	0.0435 [0.0427, 0.0442]	0.0103 [0.0099, 0.0108]	0.8645 [0.8621, 0.8670]
IID	Fully Guided Oracle	0.0228 [0.0225, 0.0232]	0.0344 [0.0339, 0.0349]	0.0074 [0.0070, 0.0078]	0.9156 [0.9148, 0.9165]
IID	Unguided	0.0285 [0.0280, 0.0290]	0.0441 [0.0435, 0.0448]	0.0114 [0.0109, 0.0120]	0.8558 [0.8534, 0.8581]
IID	Seg-Reg Cascade	0.0285 [0.0280, 0.0291]	0.0451 [0.0444, 0.0458]	0.0098 [0.0093, 0.0103]	0.8584 [0.8561, 0.8607]
IID	Mask-Sup Cascade	0.0287 [0.0281, 0.0292]	0.0456 [0.0448, 0.0463]	0.0138 [0.0131, 0.0146]	0.8601 [0.8579, 0.8623]
IID	PAPA-Net	0.0326 [0.0320, 0.0331]	0.0509 [0.0501, 0.0516]	0.0160 [0.0153, 0.0168]	0.8223 [0.8195, 0.8249]
Held-out	Direct U-Net	0.0375 [0.0372, 0.0377]	0.0534 [0.0531, 0.0537]	0.0477 [0.0466, 0.0489]	0.8389 [0.8362, 0.8416]
Held-out	Fully Guided Oracle	0.0329 [0.0327, 0.0331]	0.0448 [0.0446, 0.0451]	0.0538 [0.0528, 0.0548]	0.9039 [0.9021, 0.9055]
Held-out	Unguided	0.0377 [0.0374, 0.0380]	0.0534 [0.0531, 0.0538]	0.0450 [0.0437, 0.0463]	0.8397 [0.8366, 0.8424]
Held-out	Seg-Reg Cascade	0.0390 [0.0388, 0.0393]	0.0559 [0.0555, 0.0563]	0.0497 [0.0484, 0.0508]	0.8280 [0.8251, 0.8306]
Held-out	Mask-Sup Cascade	0.0390 [0.0388, 0.0392]	0.0559 [0.0556, 0.0562]	0.0528 [0.0516, 0.0540]	0.8342 [0.8316, 0.8369]
Held-out	PAPA-Net	0.0422 [0.0420, 0.0425]	0.0600 [0.0596, 0.0603]	0.0336 [0.0325, 0.0348]	0.7929 [0.7900, 0.7955]

Split	Method	Dice ↑	CErr ↓	Active ↑	Collapse ↓
IID	Direct U-Net	–	–	–	–
IID	Fully Guided Oracle	1.0000 [1.0000, 1.0000]	0.0024 [0.0023, 0.0025]	2.7910 [2.7680, 2.8140]	0.1741 [0.1574, 0.1914]
IID	Unguided	0.2849 [0.2815, 0.2884]	0.2427 [0.2386, 0.2466]	1.5310 [1.4950, 1.5660]	0.8209 [0.8033, 0.8390]
IID	Seg-Reg Cascade	0.8606 [0.8578, 0.8631]	0.0127 [0.0122, 0.0132]	2.7810 [2.7570, 2.8040]	0.1763 [0.1588, 0.1945]
IID	Mask-Sup Cascade	0.8606 [0.8580, 0.8632]	0.0119 [0.0114, 0.0124]	2.7770 [2.7540, 2.8010]	0.1802 [0.1618, 0.1992]
IID	PAPA-Net	0.8515 [0.8488, 0.8544]	0.0172 [0.0166, 0.0178]	2.7870 [2.7640, 2.8100]	0.1758 [0.1584, 0.1941]
Held-out	Direct U-Net	–	–	–	–
Held-out	Fully Guided Oracle	1.0000 [1.0000, 1.0000]	0.0030 [0.0029, 0.0031]	2.9990 [2.9970, 3.0000]	0.0005 [0.0000, 0.0016]
Held-out	Unguided	0.2970 [0.2960, 0.2981]	0.2011 [0.2000, 0.2024]	1.0010 [1.0000, 1.0030]	1.0000 [1.0000, 1.0000]
Held-out	Seg-Reg Cascade	0.8496 [0.8484, 0.8507]	0.0161 [0.0156, 0.0166]	3.0000 [3.0000, 3.0000]	0.0000 [0.0000, 0.0000]
Held-out	Mask-Sup Cascade	0.8496 [0.8485, 0.8508]	0.0150 [0.0146, 0.0154]	3.0000 [3.0000, 3.0000]	0.0000 [0.0000, 0.0000]
Held-out	PAPA-Net	0.8335 [0.8324, 0.8345]	0.0180 [0.0175, 0.0185]	3.0000 [3.0000, 3.0000]	0.0000 [0.0000, 0.0000]

B Extended Baselines and Ablations

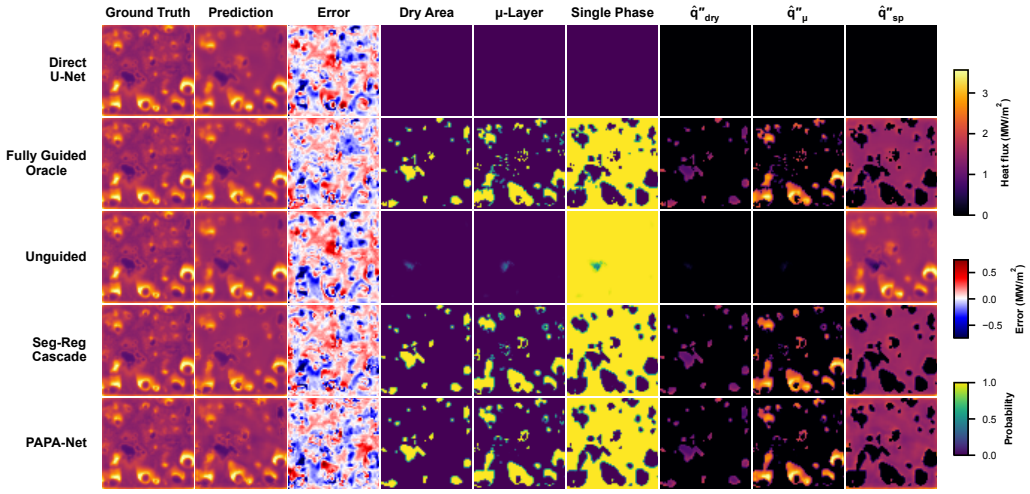


Figure 6: Qualitative mechanism diagnostics on an IID test example. This complements the held-out qualitative diagnostic in Figure 2.

Table 12: Frame-bootstrap 95% confidence intervals for mechanism-alignment ablations. Entries are point estimate [lower, upper]. Dice alone is insufficient: No Factorization maintains high mask overlap but has substantially worse contribution error, while Independent Component Heat Maps, No Phase Head, and No Teacher-Mask Guidance lose mechanism alignment or collapse.

Split	Method	Dice ↑	CErr ↓
IID	Full PAPA-Net	0.8515 [0.8490, 0.8543]	0.0172 [0.0166, 0.0177]
IID	No Factorization	0.8539 [0.8510, 0.8568]	0.0561 [0.0541, 0.0581]
IID	Independent Components	0.2797 [0.2766, 0.2827]	0.2593 [0.2552, 0.2632]
IID	No Phase Head	0.2966 [0.2942, 0.2991]	0.3274 [0.3254, 0.3294]
IID	No Teacher Guidance	0.3083 [0.3036, 0.3131]	0.1575 [0.1538, 0.1611]
Held-out	Full PAPA-Net	0.8335 [0.8323, 0.8346]	0.0180 [0.0175, 0.0184]
Held-out	No Factorization	0.8373 [0.8361, 0.8385]	0.0350 [0.0342, 0.0358]
Held-out	Independent Components	0.3376 [0.3365, 0.3387]	0.1987 [0.1974, 0.2000]
Held-out	No Phase Head	0.2836 [0.2831, 0.2842]	0.3087 [0.3075, 0.3099]
Held-out	No Teacher Guidance	0.2836 [0.2832, 0.2842]	0.2025 [0.2013, 0.2036]

Split	Method	Active ↑	Collapse ↓
IID	Full PAPA-Net	2.7870 [2.7640, 2.8090]	0.1758 [0.1588, 0.1935]
IID	No Factorization	2.7820 [2.7610, 2.8040]	0.1802 [0.1619, 0.1992]
IID	Independent Components	3.0000 [3.0000, 3.0000]	0.0000 [0.0000, 0.0000]
IID	No Phase Head	0.0000 [0.0000, 0.0000]	1.0000 [1.0000, 1.0000]
IID	No Teacher Guidance	1.1690 [1.1510, 1.1840]	1.0000 [1.0000, 1.0000]
Held-out	Full PAPA-Net	3.0000 [3.0000, 3.0000]	0.0000 [0.0000, 0.0000]
Held-out	No Factorization	3.0000 [3.0000, 3.0000]	0.0000 [0.0000, 0.0000]
Held-out	Independent Components	3.0000 [3.0000, 3.0000]	0.0000 [0.0000, 0.0000]
Held-out	No Phase Head	0.0000 [0.0000, 0.0000]	1.0000 [1.0000, 1.0000]
Held-out	No Teacher Guidance	1.0000 [1.0000, 1.0000]	1.0000 [1.0000, 1.0000]

Table 13: Frame-bootstrap 95% confidence intervals for key teacher-mask fractions. Entries are point estimate [lower, upper]. The transition from 0% to 1% teacher masks prevents collapse on the held-out condition while inference remains optical-only.

Split	Teacher masks	Dice ↑	CErr ↓	Collapse ↓	RelE ↓
IID	0%	0.3089 [0.3045, 0.3136]	0.1603 [0.1568, 0.1639]	1.0000 [1.0000, 1.0000]	0.0135 [0.0129, 0.0142]
IID	1%	0.8176 [0.8147, 0.8204]	0.0284 [0.0275, 0.0293]	0.2098 [0.1922, 0.2285]	0.0265 [0.0254, 0.0275]
IID	100%	0.8515 [0.8488, 0.8544]	0.0172 [0.0166, 0.0178]	0.1758 [0.1590, 0.1936]	0.0160 [0.0153, 0.0168]
Held-out $q'' = 1.82$	0%	0.2836 [0.2832, 0.2843]	0.2025 [0.2013, 0.2038]	1.0000 [1.0000, 1.0000]	0.0343 [0.0331, 0.0353]
Held-out $q'' = 1.82$	1%	0.8132 [0.8120, 0.8142]	0.0302 [0.0296, 0.0308]	0.0000 [0.0000, 0.0000]	0.0360 [0.0348, 0.0372]
Held-out $q'' = 1.82$	100%	0.8335 [0.8323, 0.8345]	0.0180 [0.0175, 0.0184]	0.0000 [0.0000, 0.0000]	0.0336 [0.0325, 0.0346]

B.1 Full Architecture Ablation

Table 14 reports the full architecture ablation on the IID test split and held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$) condition. Variants without explicit phase structure can fit total heat flux maps but fail to recover physically meaningful mechanism partitions.

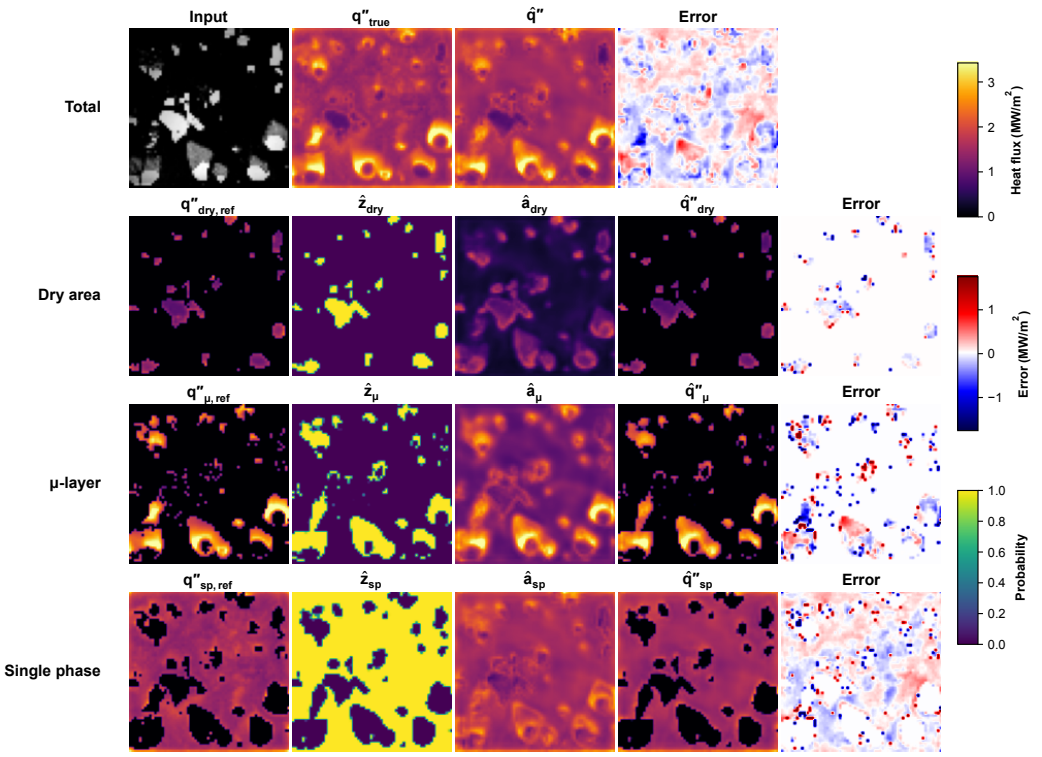


Figure 7: PAPA-Net factorization diagnostic on the IID test split. PAPA-Net receives only the optical input at inference and predicts mechanism supports $\hat{z}_k$, nonnegative amplitudes $\hat{a}_k$, and component heat flux maps $\hat{q}''_k = \hat{z}_k \hat{a}_k$. Reference components are teacher-derived as $q''_{k,\mathrm{ref}} = m_k q''$, and error maps show $\hat{q}''_k - q''_{k,\mathrm{ref}}$. This IID diagnostic complements the held-out $q'' = 1.82\ \mathrm{MW/m^2}$ diagnostic in Figure 3.

Table 14: Architecture ablation on IID and held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$). Heat flux map accuracy alone is insufficient: ablations without explicit phase structure can fit total heat flux maps but fail to recover physically meaningful mechanism partitions.

Split	Method	MAE ↓	RMSE ↓	RelE ↓	SSIM ↑	Dice ↑	CErr ↓	Active ↑	Collapse ↓
IID	Full PAPA-Net	0.0326	0.0509	0.0160	0.8223	0.8515	0.0172	2.7870	0.1760
IID	No Factorization	0.0306	0.0473	0.0190	0.8490	0.8539	0.0561	2.7820	0.1800
IID	Independent Heat Flux Maps	0.0269	0.0420	0.0106	0.8762	0.2797	0.2593	3.0000	0.0000
IID	Sigmoid Masks	0.0356	0.0559	0.0200	0.7946	0.8519	0.0283	2.8160	0.1690
IID	No Phase Head	0.0267	0.0417	0.0101	0.8785	0.2966	0.3274	0.0000	1.0000
Held-out $q'' = 1.82$	Full PAPA-Net	0.0422	0.0600	**0.0336**	0.7929	0.8335	0.0180	3.0000	0.0000
Held-out $q'' = 1.82$	No Factorization	0.0401	0.0571	0.0354	0.8136	0.8373	0.0350	3.0000	0.0000
Held-out $q'' = 1.82$	Independent Heat Flux Maps	0.0377	0.0528	0.0529	0.8510	0.3376	0.1987	3.0000	0.0000
Held-out $q'' = 1.82$	Sigmoid Masks	0.0465	0.0663	0.0403	0.7573	0.8357	0.0334	3.0000	0.0000
Held-out $q'' = 1.82$	No Phase Head	0.0384	0.0536	0.0571	0.8504	0.2836	0.3087	0.0000	1.0000

B.2 Full Loss Ablation

Table 15 reports the complete loss ablation on both the IID test split and the held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$) condition. Teacher-mask guidance is the dominant term for teacher-guided mechanism alignment: removing it preserves competitive heat flux map error but collapses the physical branches. Contribution, energy, anti-collapse, temporal, prior, and smoothness terms provide smaller stabilizing effects.

Table 15: Full loss ablation of PAPA-Net on IID and held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$) splits.

Split	Method	MAE ↓	RMSE ↓	RelE ↓	SSIM ↑	Dice ↑	CErr ↓	Active ↑	Collapse ↓
IID	Full PAPA-Net	0.0326	0.0509	0.0160	0.8223	0.8515	0.0172	2.7870	0.1758
IID	No Teacher-Mask Guidance	0.0279	0.0437	0.0129	0.8594	0.3083	0.1575	1.1690	1.0000
IID	No Contribution Teacher Loss	0.0329	0.0532	0.0194	0.7962	0.8523	0.0199	2.7860	0.1780
IID	No Energy Loss	0.0320	0.0494	0.0365	0.8272	0.8504	0.0174	2.7890	0.1735
IID	No Anti-Collapse/Decorrelation	0.0322	0.0504	0.0211	0.8263	0.8536	0.0175	2.7850	0.1741
IID	No Temporal Consistency	0.0326	0.0511	0.0172	0.8225	0.8503	0.0154	2.7870	0.1758
IID	No Area/Contribution Priors	0.0320	0.0500	0.0177	0.8388	0.8512	0.0113	2.7900	0.1735
IID	No Smoothness/TV	0.0324	0.0507	0.0179	0.8217	0.8528	0.0165	2.7860	0.1763
Held-out $q'' = 1.82$	Full PAPA-Net	0.0422	0.0600	0.0336	0.7929	0.8335	0.0180	3.0000	0.0000
Held-out $q'' = 1.82$	No Teacher-Mask Guidance	0.0364	0.0520	0.0357	0.8414	0.2836	0.2025	1.0000	1.0000
Held-out $q'' = 1.82$	No Contribution Teacher Loss	0.0421	0.0602	0.0404	0.7914	0.8352	0.0189	3.0000	0.0000
Held-out $q'' = 1.82$	No Energy Loss	0.0398	0.0566	0.0352	0.8108	0.8339	0.0177	3.0000	0.0000
Held-out $q'' = 1.82$	No Anti-Collapse/Decorrelation	0.0420	0.0599	0.0421	0.7960	0.8371	0.0203	3.0000	0.0000
Held-out $q'' = 1.82$	No Temporal Consistency	0.0423	0.0601	0.0357	0.7944	0.8337	0.0163	3.0000	0.0000
Held-out $q'' = 1.82$	No Area/Contribution Priors	0.0425	0.0604	0.0373	0.7976	0.8332	0.0135	3.0000	0.0000
Held-out $q'' = 1.82$	No Smoothness/TV	0.0420	0.0599	0.0369	0.7945	0.8354	0.0173	3.0000	0.0000

B.3 Teacher Mask Label Efficiency

Table 16 and Figure 8 report the full teacher-mask fraction sweep. Even sparse teacher supervision sharply improves mechanism alignment, especially on held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$) where collapse disappears at 1% supervision.

Table 16: Teacher-mask label efficiency on IID and held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$) splits. PAPA-Net uses teacher masks only during training and remains optical-only at inference.

Split	Teacher masks	Dice ↑	CErr ↓	Collapse ↓	RelE ↓
IID	0%	0.3089	0.1603	1.0000	0.0135
IID	1%	0.8176	0.0284	0.2098	0.0265
IID	5%	0.8281	0.0306	0.1998	0.0222
IID	10%	0.8314	0.0263	0.1987	0.0209
IID	25%	0.8404	0.0212	0.1786	0.0204
IID	50%	0.8462	0.0196	0.1735	0.0206
IID	100%	0.8515	0.0172	0.1758	0.0160
Held-out $q'' = 1.82$	0%	0.2836	0.2025	1.0000	0.0343
Held-out $q'' = 1.82$	1%	0.8132	0.0302	0.0000	0.0360
Held-out $q'' = 1.82$	5%	0.8231	0.0331	0.0000	0.0346
Held-out $q'' = 1.82$	10%	0.8247	0.0260	0.0000	0.0337
Held-out $q'' = 1.82$	25%	0.8294	0.0222	0.0000	0.0402
Held-out $q'' = 1.82$	50%	0.8321	0.0211	0.0000	0.0372
Held-out $q'' = 1.82$	100%	0.8335	0.0180	0.0000	0.0336

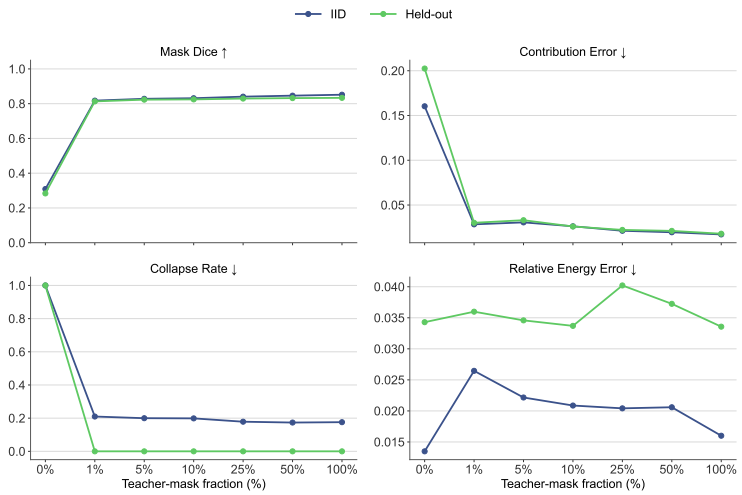


Figure 8: Teacher-mask label efficiency. PAPA-Net is trained with different fractions of teacher-mask supervision while inference remains optical-only. The transition from 0% to 1% masks prevents collapse and recovers most of the mechanism-aligned dry area, μ-layer, and single phase decomposition quality.

B.4 Full physics-aware metrics

Table 17 reports the complete physics-aware metric set summarized in Figure 9. PAPA-Net improves integrated heat balance and heat-trend consistency among optical-only methods, while the fully guided oracle remains an upper bound for mask-at-inference partition quality.

Table 17: Physics-aware evaluation on held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$). PAPA-Net improves integrated heat balance and heat-trend consistency among optical-only methods.

Method	IntE ↓	ContribErr ↓	Top 10% Heat MAE ↓	Grad 10% MAE ↓	Temporal VarErr ↓	Trend Corr ↑
Direct U-Net	0.0477	—	0.0838	0.0669	—	0.2877
Seg-Reg Cascade	0.0497	0.0161	0.0837	0.0715	0.0098	0.3815
Mask-Sup Cascade	0.0528	0.0150	0.0826	0.0703	0.0098	0.4266
PAPA-Net	**0.0336**	0.0180	0.0895	0.0744	0.0097	**0.4577**
Fully Guided Oracle	0.0538	**0.0030**	**0.0623**	**0.0556**	**0.0025**	0.3148

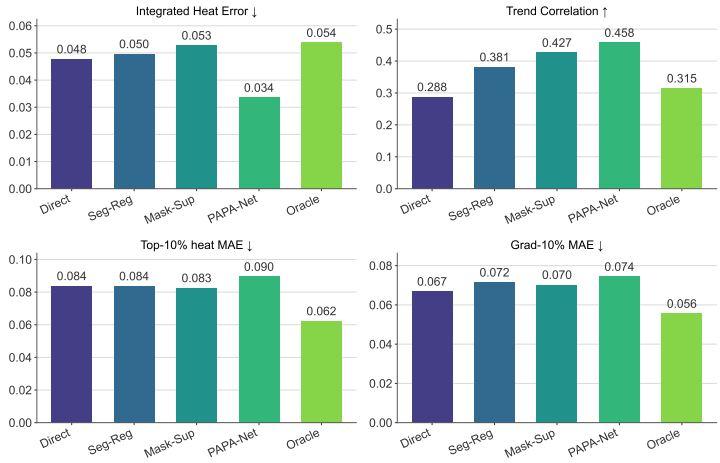


Figure 9: Physics-aware evaluation on held-out $q'' = 1.82\ \mathrm{MW/m^2}$ ($q''/q''_{\mathrm{CHF}} = 0.83$). PAPA-Net has the lowest integrated heat error and highest trend correlation among optical-only methods, with competitive high-heat and high-gradient error. Oracle uses masks at inference.

B.5 Robustness to Optical and Teacher Mask Corruptions

Figure 10 evaluates two perturbation settings: inference-time optical corruptions and training-time teacher-mask corruptions. PAPA-Net is stable under additive Gaussian noise and brightness/contrast shifts, indicating that the learned factorization is not overly sensitive to mild image-level noise. Blur has a larger effect because it removes local phase boundary information needed to distinguish dry area, μ-layer, and single phase regions. Teacher-mask corruption produces gradual degradation in Dice and contribution error, with severe perturbations causing the largest loss of mechanism localization. Overall, the model is robust to mild experimental noise, but accurate phase boundary information remains important for mechanism-resolved heat flux partitioning.

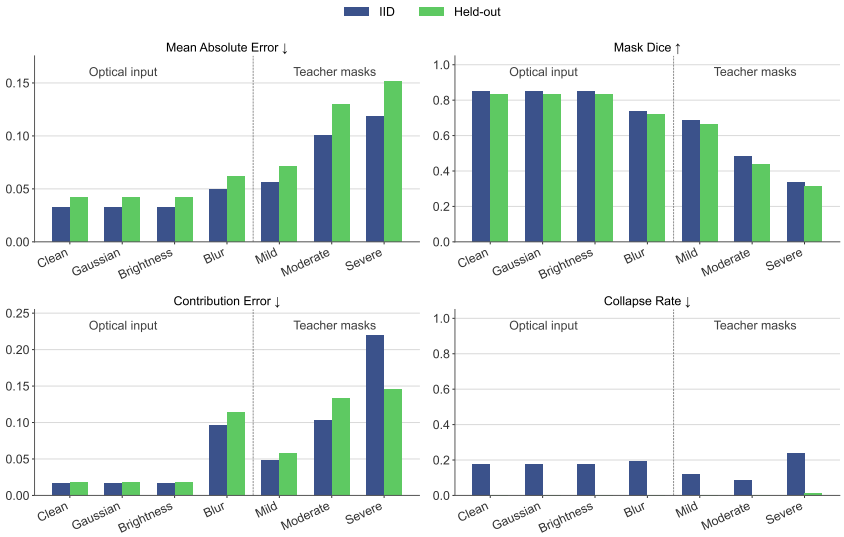


Figure 10: Robustness to optical corruptions and imperfect teacher masks. Mild noise and brightness/contrast shifts have little effect; blur and severe mask perturbations degrade mechanism recovery. Inference remains optical-only.

B.6 Leave One Condition Out Robustness and Low-Flux Outlier Audit

Table 18 reports leave-one-condition-out robustness across applied heat flux conditions. Source sequence IDs are shown only for traceability. PAPA-Net maintains active, noncollapsed partitions across most held-out applied heat flux conditions, especially the mid-to-high heat flux regimes. The main outlier is the low-flux condition $q'' = 0.493\ \mathrm{MW/m^2}$ (source S4), which exhibits large relative energy error and partial component collapse.

Table 18: Leave-one-condition-out robustness across applied heat flux conditions. Source IDs are included only for traceability. The low-flux condition $q'' = 0.493\ \mathrm{MW/m^2}$ (source S4) is the hardest held-out regime, while most other conditions preserve active, noncollapsed physical branches.

Held-out condition	MAE ↓	RMSE ↓	RelE ↓	SSIM ↑	Dice ↑	CErr ↓	Active ↑	Collapse ↓
$q'' = 0.493$ (S4)	0.1101	0.1157	1.1200	0.5637	0.7870	0.0090	1.9870	0.9260
$q'' = 1.138$ (S6)	0.0649	0.0768	0.2587	0.7943	0.8595	0.0104	2.9270	0.0730
$q'' = 1.529$ (S7)	0.0431	0.0580	0.0781	0.8167	0.8455	0.0102	3.0000	0.0000
$q'' = 1.684$ (S8)	0.0411	0.0567	0.0406	0.8125	0.8425	0.0205	3.0000	0.0000
$q'' = 1.823$ (S9)	0.0421	0.0600	0.0365	0.7940	0.8322	0.0170	3.0000	0.0000
$q'' = 1.941$ (S11)	0.0497	0.0699	0.0535	0.7575	0.8144	0.0295	3.0000	0.0000
$q'' = 2.172$ (S12)	0.0699	0.0959	0.1241	0.6622	0.7735	0.0328	3.0000	0.0000

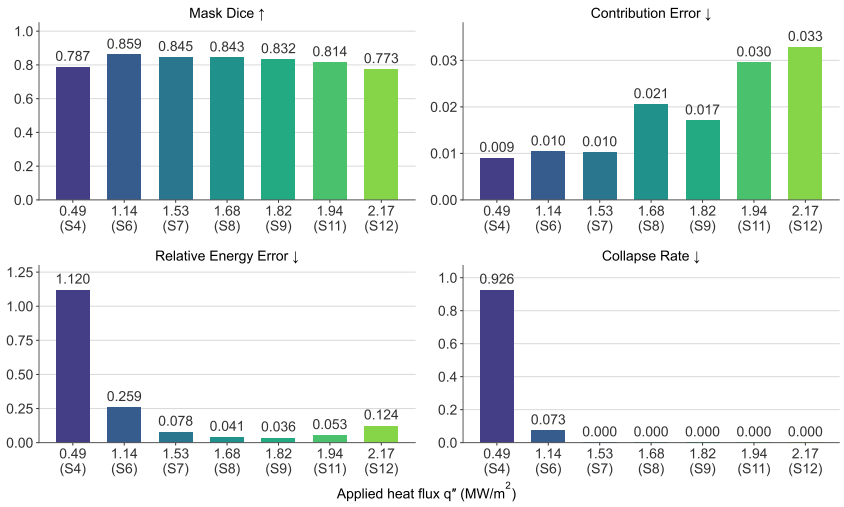


Figure 11: Leave-one-condition-out robustness across applied heat flux conditions. Each tick shows the held-out applied heat flux q'' in $\mathrm{MW/m^2}$, with the source ID in parentheses. PAPA-Net preserves noncollapsed partitions across most held-out conditions, while $q'' = 0.493\ \mathrm{MW/m^2}$ (source S4) is a low-energy, near-single phase outlier.

Table 19 diagnoses why the low-flux condition $q'' = 0.493\ \mathrm{MW/m^2}$ (source S4) is difficult. This condition has substantially lower true heat energy and is nearly single phase dominated, with very rare dry area activity. It therefore underexcites the dry area and μ-layer branches, explaining the high relative energy error and partial collapse.

Table 20: dT10-Boiling evaluated heat-flux steps.

Step	Applied heat flux q'' (MW/m^2)
4	0.3506
6	0.5464
10	0.9456
12	1.1469
14	1.3485
16	1.5473
18	1.7473
20	1.9419

Table 19: Low-flux outlier audit for $q'' = 0.493\ \mathrm{MW/m^2}$ (source S4). This condition is low-energy and near-single phase, with rare dry area events.

Metric	$q'' = 0.493$ (S4)	Other median	Ratio
AbsE	449.7	91.29	4.9260
RelE	1.120	0.0658	17.0200
True energy mean	402.1	1434	0.2803
True energy min	373.3	1310	0.2850
Dice	0.7870	0.8374	0.9399
Active components	1.987	3.000	0.6622
Collapse	0.926	0.000	—
Teacher dry area	0.00281	0.08469	0.0332
Teacher μ area	0.01971	0.1673	0.1178
Teacher single phase area	0.9775	0.7480	1.3070
Dry event rate	0.0285	1.000	0.0285
μ event rate	0.8230	1.000	0.8230
Single phase event rate	1.000	1.000	1.0000

B.7 dT10-Boiling zero-shot transfer

We evaluate zero-shot transfer on dT10-Boiling, an external flow-boiling dataset acquired under changed thermal and acquisition conditions. The dataset uses deionized-water upflow at pressure 1.21 bar, mass flux $500\ \mathrm{kg/m^2/s}$, channel size $3\ \mathrm{cm} \times 1\ \mathrm{cm}$, and $\Delta T = 10^{\circ}\mathrm{C}$ subcooling. Videos were recorded at 3000 FPS for approximately 2 s, with 31 μm/px spatial resolution after alignment to the IR coordinate system; the heater area is $98.6049\ \mathrm{mm^2}$. Raw files contain optical phase-detection videos, heat flux to sapphire, temperature maps, and binary phase masks. We convert each valid step into the same HDF5 interface used by PAPA-Net, producing optical input, heat target, and teacher-mask files with shape $6000 \times 230 \times 283$. The water-side heat target is computed as the applied heat flux minus the aligned sapphire heat-flux map. Dry masks are taken from the binary dry mask; μ-layer masks are the evaporation mask excluding dry pixels; and single phase masks are the complement. Existing checkpoints are evaluated without retraining, fine-tuning, calibration, or model selection on the external data.

The aggregate external metrics are reported in Table 1; Figure 12 shows per-step trends.

C Runtime and Deployment Analysis

Table 21 reports inference cost at batch size 1. PAPA-Net has nearly the same parameter count and runtime as Direct U-Net, while producing mechanism-resolved heat flux components. It is also substantially lighter than the cascade baselines.

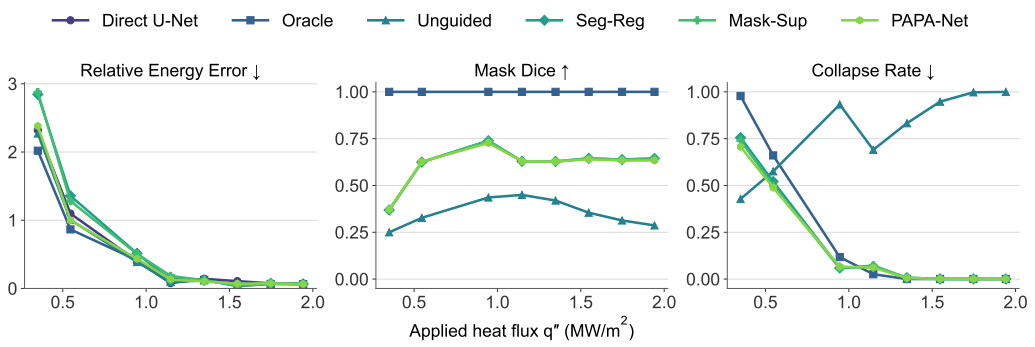


Figure 12: Per-step zero-shot transfer to dT10-Boiling under changed thermal/acquisition conditions. Existing checkpoints are evaluated without retraining. PAPA-Net preserves noncollapsed mechanism partitions in the mid/high heat-flux regimes, while low-flux cases remain difficult for all methods because relative integrated error is amplified when total heat is small.

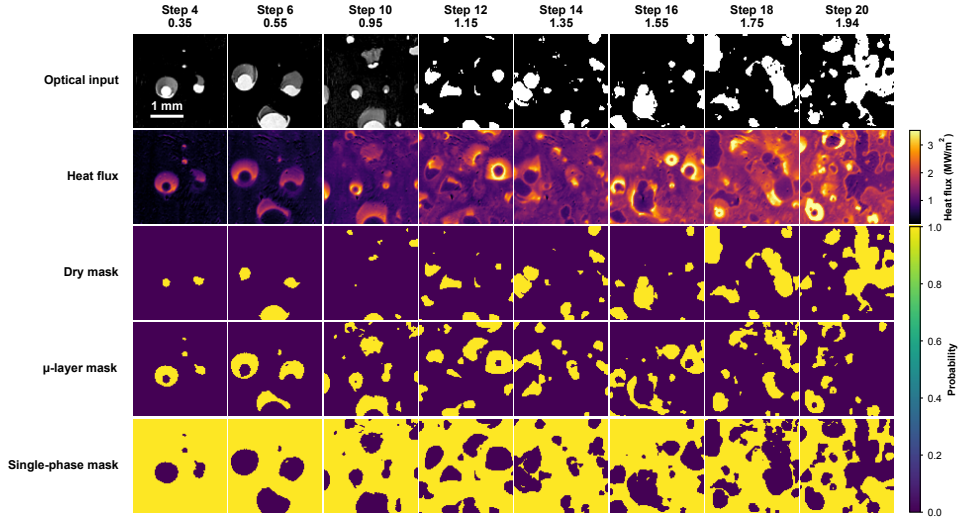


Figure 13: dT10-Boiling dataset overview across applied heat-flux steps. Rows show optical phase-detection input, IR heat-flux map, and teacher masks for dry area, μ-layer, and single phase. Teacher masks are used only for evaluation diagnostics; PAPA-Net inference uses only optical input. The scale bar denotes 1 mm.

Table 21: Runtime and deployment analysis at batch size 1. Fully Guided Oracle uses masks at inference and is included only as an oracle reference.

Method	Params ↓	GPU memory MB ↓	ms/frame ↓	FPS ↑	Inference input
Direct U-Net	5.30M	46.43	0.90	1109	optical-only
Fully Guided Oracle	5.31M	46.89	0.97	1034	optical + masks
Seg-Reg Cascade	10.59M	88.52	2.09	479	optical-only
Mask-Sup Cascade	10.59M	88.21	2.09	480	optical-only
PAPA-Net	5.31M	48.51	0.97	1029	optical-only

D Controlled Synthetic Partition Benchmark

We additionally evaluate PAPA-Net on a controlled synthetic partition benchmark with known mechanism masks and nonnegative component amplitudes. This benchmark is not intended to replace the real boiling data; instead, it isolates the core mechanism-alignment question under known ground

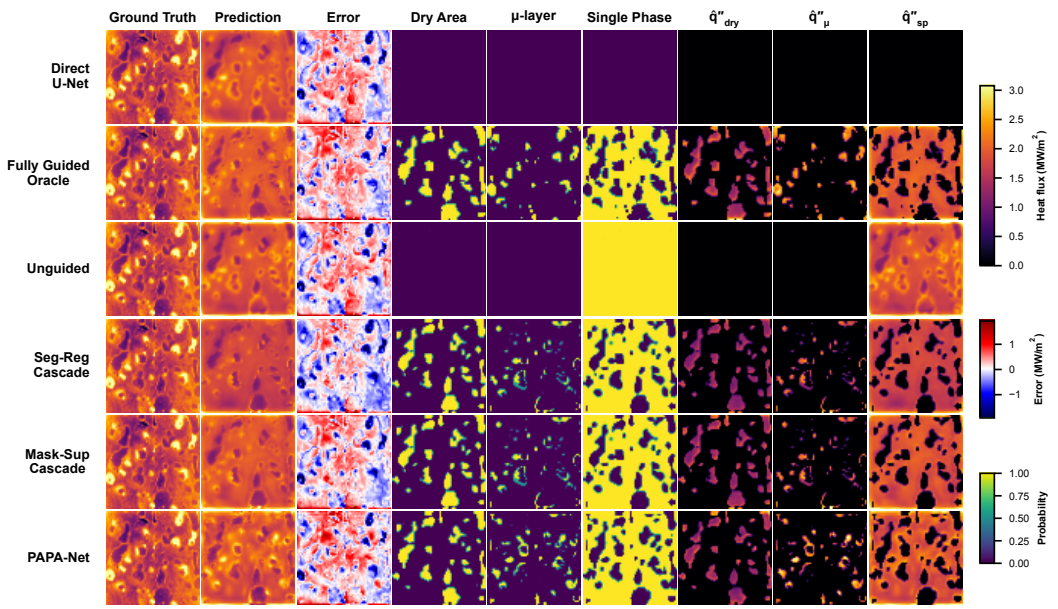


Figure 14: Qualitative zero-shot dT10-Boiling mechanism diagnostics. Direct U-Net predicts only total heat flux; Fully Guided Oracle uses masks at inference; Unguided decomposition collapses or misassigns branches; and PAPA-Net remains optical-only while producing dry area, μ-layer, and single phase heat-flux components.

truth: can a model recover physically named component masks and heat contributions, rather than merely fitting the total heat flux map?

Because the synthetic data are generated from known mask-amplitude factors, this benchmark is intended as a controlled mechanism-alignment stress test rather than a substitute for independent real-boiling validation.

D.1 Synthetic data generation

Each synthetic sample contains an optical input $x \in [0,1]^{64\times64\times3}$, a total heat flux map q'', and ground-truth mechanism masks for dry area, μ-layer, and single phase. Dry regions are generated as unions of randomly rotated ellipses. Let z_{dry} denote this binary dry mask. The μ-layer is constructed as a narrow ring around dry regions:

$$z_\mu = \mathrm{dilate}(z_{\mathrm{dry}}, r) \setminus \mathrm{erode}(z_{\mathrm{dry}}, 1), \tag{33}$$

and the single phase region is the complement

$$z_{\mathrm{sp}} = 1 - (z_{\mathrm{dry}} \vee z_\mu)\,. \tag{34}$$

For IID samples, ellipses have radii in $[4, 12]$ pixels and a fixed μ-layer dilation width $r = 2$. For OOD samples, ellipses can be larger, with radii up to 18 pixels, the number of dry blobs increases, and r is sampled from $\{2, 3, 4\}$.

We then generate smooth mechanism-specific amplitude fields. Let T denote a normalized smooth random texture and let U_s denote a random field smoothed with a box filter of width s. The IID amplitudes are

$$a_{\mathrm{dry}} = 0.6 + 0.5T, \qquad a_\mu = 1.8 + U_3, \qquad a_{\mathrm{sp}} = 1.0 + 0.4U_7. \tag{35}$$

For OOD samples, amplitudes are shifted as

$$a_{\mathrm{dry}} \leftarrow 0.8a_{\mathrm{dry}}, \qquad a_\mu \leftarrow 1.3a_\mu, \qquad a_{\mathrm{sp}} \leftarrow 1.1a_{\mathrm{sp}}. \tag{36}$$

The component and total heat flux maps are exactly

$$q''_k = z_k a_k, \qquad q'' = \sum_{k\in\{\mathrm{dry},\mu,\mathrm{sp}\}} q''_k. \tag{37}$$

The optical input is a noisy three channel proxy derived from phase, boundary, texture, and heat information:

$$x_1 = 0.25 z_{\mathrm{dry}} + 0.65 z_\mu + 0.35 z_{\mathrm{sp}} + 0.15 T, \tag{38}$$

$$x_2 = \mathrm{blur}(z_\mu, 3) + \eta, \tag{39}$$

$$x_3 = 0.3\, q'' / \max(q'') + 0.7 T. \tag{40}$$

Gaussian noise is added to all channels and the result is clipped to $[0, 1]$. The OOD split uses a larger noise level than the IID splits.

D.2 Splits and purpose

The benchmark contains 3000 training samples, 500 validation samples, 500 IID test samples, and 1000 OOD samples, totaling 5000 samples. The OOD split changes geometry, μ-layer thickness, amplitude scaling, and optical noise. This benchmark directly probes mechanism alignment under known factors: Direct U-Net can fit q'' but has no partition output, unguided independent components can reproduce plausible heat-flux patterns but fail to recover mechanism identity, while PAPA-Net should recover the known dry-area, μ-layer, and single phase partitions using optical input only at inference.

Table 22: Controlled synthetic partition benchmark. The IID split tests interpolation under the same procedural distribution, while the OOD split changes geometry, μ-layer thickness, amplitude scaling, and optical noise. PAPA-Net recovers near oracle mechanism partitions while remaining optical-only at inference.

Split	Method	MAE ↓	RelE ↓	Dice ↑	CErr ↓	Active ↑	Collapse ↓
Synthetic IID	Direct U-Net	0.0406	0.0023	—	—	—	—
Synthetic IID	Unguided	0.0381	0.0021	0.2713	0.1886	0.0000	1.0000
Synthetic IID	Seg-Reg Cascade	0.0409	0.0019	0.9999	0.0018	3.0000	0.0000
Synthetic IID	PAPA-Net	0.0426	0.0026	0.9999	0.0013	3.0000	0.0000
Synthetic IID	Fully Guided Oracle	0.0372	0.0019	1.0000	0.0012	3.0000	0.0000
Synthetic OOD	Direct U-Net	0.2880	0.1062	—	—	—	—
Synthetic OOD	Unguided	0.2766	0.1023	0.1964	0.1581	0.0000	1.0000
Synthetic OOD	Seg-Reg Cascade	0.3438	0.1270	0.9956	0.0607	3.0000	0.0000
Synthetic OOD	PAPA-Net	0.3040	0.1260	0.9973	0.0395	3.0000	0.0000
Synthetic OOD	Fully Guided Oracle	0.3006	0.1157	1.0000	0.0381	3.0000	0.0000

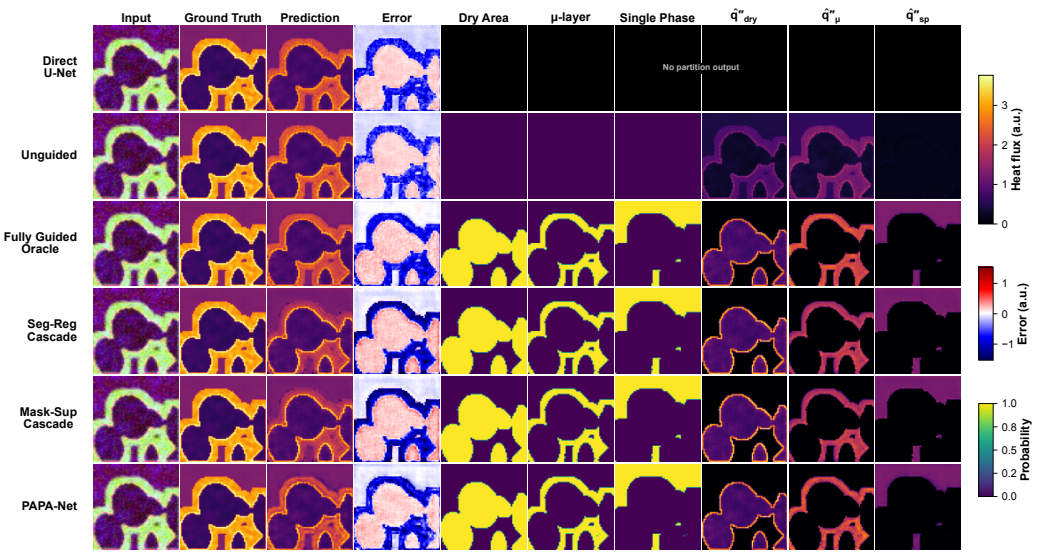


Figure 15: Qualitative results on the controlled synthetic partition benchmark. Ground-truth components are known by construction, and synthetic heat-flux fields are shown in arbitrary units (a.u.). Direct U-Net predicts only the total heat-flux map and provides no mechanism decomposition. Unguided and cascade baselines produce plausible heat-flux patterns but do not recover mechanism partitions as reliably. PAPA-Net recovers dry-area, μ-layer, and single-phase masks together with corresponding heat-flux components while using only the synthetic optical input at inference.