Practical Advances in Artificial Intelligence and Computing
Frontier Research Publishing • https://www.frpubs.org
Vol. 4, No. 4, 2025

Temporal Evidence Accumulation Trades Alarm Purity Against Detection Delay in Synthetic Melt Pool Monitoring

Veylora Thistmere[1]*, Caedric Osenvale[1], Nyssara Brindleth[1]
[1]University of Manitoba, Winnipeg, Canada
*Corresponding author: veylorathistmere684@gmail.com

Abstract: This fully synthetic study separates spatial frame classification from temporal evidence accumulation in melt pool monitoring. A seeded generator creates 9,600 radiance images across 150 independent sixty four frame tracks, including sustained morphology changes, track offsets and isolated spatter flashes. Penalized logistic regression learns a score from five explicit image features using sixty training tracks; twenty validation tracks determine score thresholds, and seventy test tracks provide evaluation. A two state Bayesian filter converts classifier probabilities to likelihood ratios and accumulates causal evidence. On 4,480 test frames, frame logistic F1 is 0.339, nearly identical to thresholded brightness at 0.341. Temporal F1 increases to 0.604 while false positive frames fall from 126 to 96. However, detected events decrease from sixty to fifty seven of sixty three, and mean delay rises from 2.42 to 3.65 frames. Permuting frame order reduces temporal F1 to 0.211. The results demonstrate an explicit purity versus delay tradeoff for generated morphology events, not defect detection in manufactured parts.

Keywords: melt pool monitoring; temporal filtering; logistic regression; event detection delay; synthetic radiance; additive manufacturing

1 A bright frame is not a persistent process event

A melt pool camera records a succession of luminous shapes rather than a direct declaration that a manufacturing defect has occurred. A short spatter flash, an elongated pool and a sustained change in the process can all increase image brightness. Their operational meanings differ because duration matters. This paper examines an inspectable alternative to treating every frame as an independent classification decision. We first learn a frame score from five image measurements, then accumulate that score through a two state Bayesian filter. The question is whether persistence improves alarm quality and what it costs in event detection delay.

The scope is deliberately narrower than identifying porosity in a manufactured component. High speed X ray observations have revealed keyhole thresholds and complex pore formation mechanisms [1] [2]. Real time imaging and diffraction also show the richness of signals available during laser powder bed fusion [3]. Such observations establish why process monitoring is scientifically valuable. They do not make a bright optical feature equivalent to a pore, nor do they supply labels for the numerical experiment reported here. Our hidden event means a sustained change in synthetic image morphology, not a verified defect in metal.

Recent primary studies propose learned feature embeddings, adaptable segmentation and engineering guided representations for melt pool monitoring [4] [5] [6]. Their contributions motivate careful representation design. Our experiment instead keeps the spatial model small enough to reveal the separate contribution of temporal evidence. The executed learner is penalized logistic regression on mean radiance, hot pixel fraction, two second moments and peak radiance. No convolutional network, segment anything

checkpoint or vision transformer is trained. Those published methods are research context, not architectures whose benchmark performance is being reproduced.

The principal result is a tradeoff. On 4,480 held-out test frames from seventy independently generated tracks, temporal accumulation improves frame F1 from 0.339 for the learned frame classifier to 0.604 and reduces false positive frames from 126 to 96. It also increases mean delay among detected events from 2.42 to 3.65 frames and detects fewer of the sixty three generated events. A shuffled order control reduces temporal F1 to 0.211. Thus temporal persistence is useful evidence in this model, but it cannot be described as a free improvement on every monitoring objective.

Figure 1 is an AI generated illustrative laser powder bed fusion scene. Its bright interaction zone and spatter supply domain context only. The numerical frames are mathematical radiance fields generated on a thirty two by forty eight grid. Neither the scene nor the numerical images are acquired camera measurements, thermal calibration data, actual manufactured parts or public benchmark samples. Keeping these sources distinct allows a realistic opening image without creating an unsupported claim of laboratory activity or component quality validation.

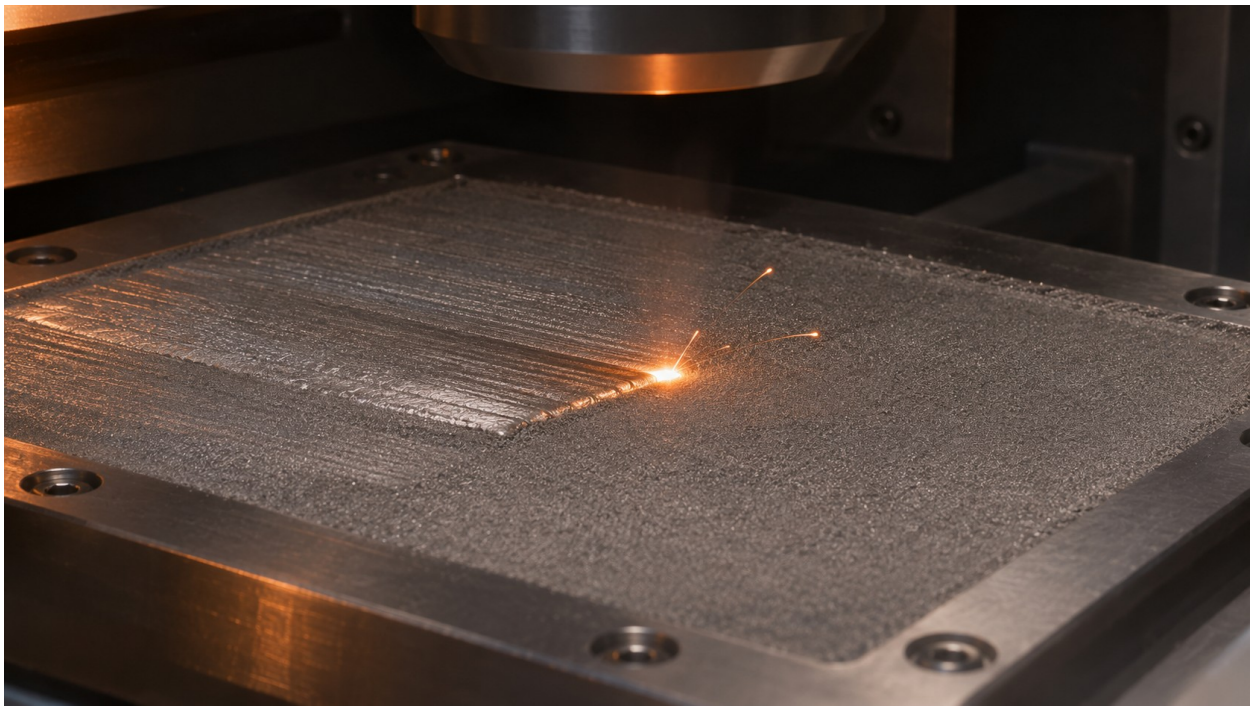


Figure 1. AI-generated illustrative laser powder bed fusion process scene. The bright pool and spatter are visual context only, not acquired images, temperature measurements or evidence of part quality.

2 Constructing events and observations separately

Every synthetic track contains sixty four frames. With probability 0.85, the generator inserts one sustained event whose starting frame is uniformly sampled from fifteen through forty two and whose duration is an integer from seven through fourteen frames. The remaining tracks contain no event. Track membership defines the independent unit for splitting training, validation and test data. There are sixty training tracks, twenty validation tracks and seventy test tracks. This split prevents frames immediately adjacent to a test event from appearing in training merely because images were shuffled at the frame level.

$$z_t \in \{0,1\}, d \in \{7,8,\ldots,14\} \tag{1}$$

Within a frame, an event shifts a latent morphology variable by 0.75. Independent normal fluctuation with standard deviation 0.60 is added at every time step. That fluctuation is large enough to make individual event and background images overlap. The event does not deterministically create a perfectly separable bright blob. A small track specific radiance offset adds a repeated background difference across frames. The purpose is to give the temporal model a real inference problem: it must combine weak observations over time rather than simply apply a threshold to an almost noiseless signal.

$$a_t = 0.75 z_t + \eta_t, \eta_t \sim N(0, 0.60^2) \tag{2}$$

The central pool is an elliptical Gaussian, with axial and transverse widths changed exponentially by the latent variable. Its amplitude changes more weakly, and a faint downstream tail is added. Coordinates span three arbitrary axial units and two transverse units. The Gaussian construction captures elongation, transverse spreading and radiance changes in a transparent way. It does not solve a heat equation or a free surface fluid model. Spatial units, radiance units and time steps are numerical coordinates, so one frame of delay cannot be converted into microseconds or scan distance without additional assumptions.

$$L_t = 0.31 exp(0.22 a_t), W_t = 0.19 exp(0.10 a_t) \tag{3}$$

$$I_t(x, y) = b + A_t exp(-(x + 0.12)^2 / L_t^2 - y^2 / W_t^2) + q(x, y) \tag{4}$$

Spatter is generated independently of the sustained event. At each frame there is a 0.12 probability of a narrow off center Gaussian flash with random location and amplitude. Independent pixel noise with standard deviation 0.018 is then added and negative radiance is clipped to zero. This flash mechanism creates bright maxima without necessarily increasing the broad hot area in the same way as an elongated central pool. It is a useful nuisance because a detector that responds only to the brightest pixel should confuse some flashes with meaningful morphology changes.

Table 1. Synthetic image generation and split specification. No acquired manufacturing or quality data are used.

Quantity	Executed value	Interpretation
Grid; track length	32 × 48; 64 frames	Radiance and coordinates arbitrary
Training; validation; test	60; 20; 70 tracks	Track independent splits
Latent event shift; noise	0.75; standard deviation 0.60	Substantial frame overlap
Spatter probability	0.12 per frame	Independent narrow radiance flash
Event duration; master seed	7–14 frames; 2025110102	At most one event per track

The construction draws on process distinctions rather than pretending to reproduce all manufacturing physics. Keyhole mode melting and complex melt flow simulations show that optical appearance can reflect several interacting mechanisms [7] [8]. Denudation studies further demonstrate that powder motion around the pool has its own spatial dynamics [9]. We omit laser power, scan speed, absorptivity, material properties and gas flow as explicit variables. The generator is therefore a counterexample laboratory for monitoring logic, not a calibrated simulator of a particular alloy, machine or deposition recipe.

Figure 2 provides six executed numerical frames around one event. The panels show substantial overlap between states and reveal why a single image can be ambiguous. The displayed example is chosen as the first

test track containing an event, not selected for an unusually favorable detector outcome. Its quantitative interpretation remains subordinate to the full test ledger. A generated process photograph cannot provide this kind of numerical evidence, which is why Figure 1 and the computed image plate have explicitly different captions and provenance.

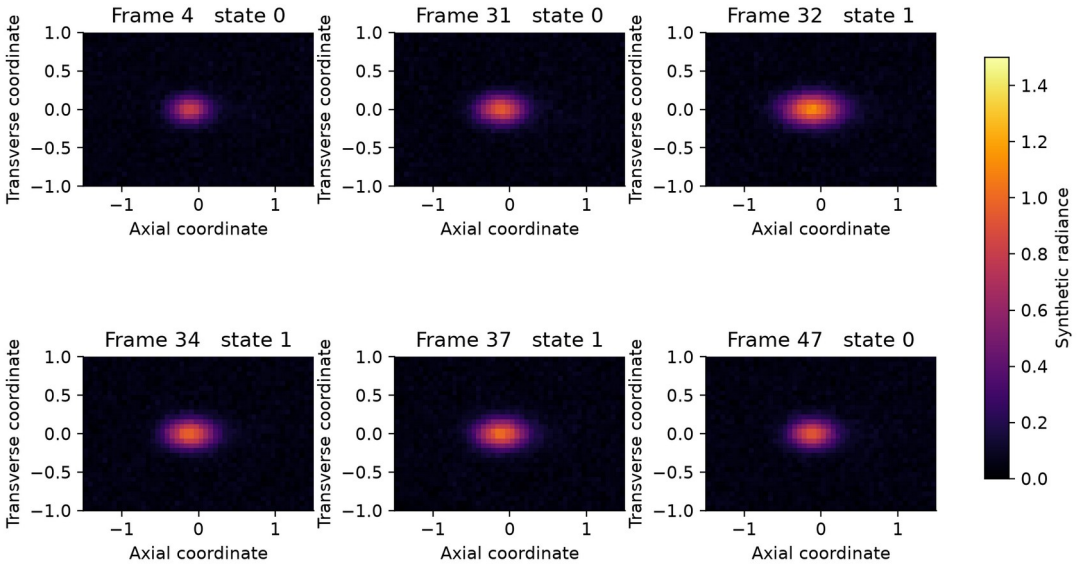


Figure 2. Executed synthetic radiance plate from the first test track containing an event. Six frames span background, onset, sustained change and recovery; coordinates and radiance have arbitrary numerical units.

3 Five measurements and one learned frame score

The five features are calculated directly from every generated image. Mean radiance measures overall brightness. Hot fraction is the proportion of pixels above 0.30 radiance units. To obtain shape moments, the program subtracts 0.10 from radiance and clips the remainder at zero, then normalizes the resulting positive weights. The weighted centroid and axial and transverse second moments summarize the spread of luminous material. The final feature is the maximum pixel value. These definitions are fixed before learning and never use the hidden event label during feature extraction.

$$w_{ij} = max(I_{ij} - 0.10, 0), m = \sum_{ij} w_{ij} \tag{5}$$

$$s_x^2 = m^{-1} \sum_{ij} w_{ij} (x_j - \bar{x})^2 \tag{6}$$

The moments are radiance weighted, not binary segmentation measurements of a metallurgical boundary. A narrow bright spatter point can move the centroid and inflate a second moment. This is not an implementation error; it reflects the fact that a simple image measurement contains both useful process information and nuisance effects. Threshold segmentation and edge detection provide established ways to extract structure [10] [11], but neither is executed here. In particular, the fixed hot fraction threshold should not be described as Otsu optimization or as an edge based melt pool contour.

All features are standardized using means and standard deviations from training tracks only. The classifier contains a constant and five feature coefficients. It minimizes binary negative log likelihood with a fixed

quadratic penalty of 0.4 on nonconstant coefficients. SciPy BFGS performs the numerical optimization. The learned coefficient on standardized hot fraction is approximately 1.067, whereas the coefficient on maximum radiance is approximately negative 0.159. These coefficients are descriptive of correlated numerical features and should not be interpreted as causal effects of pool geometry or as transferable physical laws.

$$p_t=\left(1+exp\left(-\beta^T\varphi_t\right)\right)^{-1} \tag{7}$$

$$J\left(\beta\right)=-\sum_t\left[z_t log\, p_t+\left(1-z_t\right)log\left(1-p_t\right)\right]+0.4\|\beta_{1:5}\|^2 \tag{8}$$

Ridge style penalization stabilizes correlated predictors [12]. More expressive trees and learned representations offer other options [13], while computer vision studies have directly trained anomaly classifiers for laser powder bed fusion [14]. Our modest classifier has a different purpose: it provides a reproducible probabilistic score whose temporal treatment can be changed without also changing spatial model capacity. If a complex representation were introduced simultaneously with filtering, an improvement could not be attributed cleanly to evidence accumulation rather than to the new feature extractor.

Recent multi modal data pipelines and multi signal fusion studies highlight the importance of registering process signals and organizing large monitoring datasets [15] [16]. This experiment contains a single synthetic radiance modality. Track and frame indices are exact, so there is no sensor synchronization error. The result consequently speaks to temporal inference after alignment has already been solved. It cannot establish that camera timestamps, machine coordinates or post build inspection labels can be aligned with the same accuracy in an actual additive manufacturing workflow.

4 Turning frame scores into sequential evidence

The temporal model has two states: ordinary morphology and sustained changed morphology. Its transition probability from ordinary to changed is 0.018 per frame; persistence of the changed state is 0.88. These values are fixed modeling choices, not maximum likelihood estimates of the generator's exact duration distribution. In fact, the generator inserts at most one finite event per track, whereas a stationary two state chain permits multiple events and geometrically distributed durations. The filter is intentionally a useful approximation whose adequacy is assessed by held out outcomes rather than assumed from model identity.

$$q_{t|t-1}=0.018\left(1-q_{t-1}\right)+0.88q_{t-1} \tag{9}$$

The classifier output is a posterior score under the training class frequency, which is approximately 0.15156. It cannot be inserted directly as an observation likelihood without adjusting for that prior. The program converts the score to a likelihood ratio by dividing its odds by the training prior odds. This adjustment distinguishes accumulated evidence from repeated application of the same prior information. Without it, even an uninformative classifier could steadily pull the temporal posterior toward an unintended state probability merely because the base event frequency had been counted more than once.

$$\Lambda_t=p_t\left(1-\pi\right)/\left(\left(1-p_t\right)\pi\right) \tag{10}$$

$$q_t=q_{t|t-1}\Lambda_t/\left(1-q_{t|t-1}+q_{t|t-1}\Lambda_t\right) \tag{11}$$

The filter is initialized at the training event frequency at the beginning of each track. It uses only the current and previous frame scores, so no future test image can accelerate the reported alarm. This is causal filtering rather than a backward smoother. State probability updates are a standard consequence of Bayes' rule and a transition model. Classical sequential testing and cumulative inspection established the broader value of accumulating weak evidence over time [17] [18]. Hidden state parameter learning supplies a separate historical foundation [19], but no expectation maximization training is performed here.

A Kalman filter would be appropriate for a different observation and state formulation [20]. Our hidden state is discrete and the frame classifier supplies a likelihood ratio rather than a linear Gaussian temperature measurement. Calling this model a Kalman filter would therefore be inaccurate. Likewise, the persistence probability is not a measured defect survival probability. It encodes a prior belief that a sustained event is more likely to continue than to appear in an unrelated isolated frame. The shuffled order experiment below tests the empirical value of that temporal belief.

The likelihood ratio adjustment has an intuitive interpretation. A frame probability above the training prior provides positive evidence for the changed state even when it is below one half. A probability equal to the prior provides a likelihood ratio of one and leaves the prediction step unmodified by the observation. A probability below the prior supplies negative evidence. This is why a sequence of moderately elevated scores can produce an alarm although few individual scores exceed the independent frame threshold. The state prediction and the observation update play different roles and should not be collapsed into an arbitrary moving average of probabilities.

The transition model also limits how rapidly evidence decays. Its changed state persistence of 0.88 corresponds to an expected geometric run length of about 8.33 frames in the model. The generator uses durations from seven through fourteen, whose average is 10.5. These values are of similar order but not identical. The ordinary to changed transition probability implies a stationary event frequency different from the finite track generator. Initializing at the training frequency further distinguishes the filter from a fully stationary chain. These mismatches are disclosed because a model fitted exactly to its own generating mechanism would make the comparison artificially easy.

Neither the state transition parameters nor the classifier penalty are tuned on test performance. They were specified as compact design choices before the final metrics were calculated. The validation set determines only the alarm thresholds. This avoids a subtle form of test reuse in which a temporal model appears to improve delay simply because its persistence is repeatedly adjusted after inspecting the same episodes. It does not establish that the chosen constants are broadly robust. A future selection study would need a larger validation protocol and a separate untouched evaluation set, especially if several duration models were considered.

The filter assumes that frame evidence is conditionally independent given the discrete state. The generator's per frame latent fluctuation and pixel noise are independent, but the track specific radiance offset creates some dependence beyond the state. Consequently the observation likelihood approximation is imperfect even before accounting for logistic regression error. A filter can become too confident when repeated evidence contains an unmodeled shared component. We do not attach calibrated probability claims to the posterior; it is evaluated as a monitoring score. Its numerical probability interpretation is conditional on modeling assumptions that are only approximately satisfied.

Figure 3 overlays learned frame scores, temporal posterior and the generating event interval for that same track. The posterior rises only after several informative observations and can persist briefly when an individual frame becomes less distinctive. This behavior explains both increased within event recall and delayed initial recognition. It also shows why an alarm trace must be assessed causally: drawing a smooth curve through an entire sequence can look attractive while concealing use of future observations. The present update proceeds left to right and never uses a backward pass.

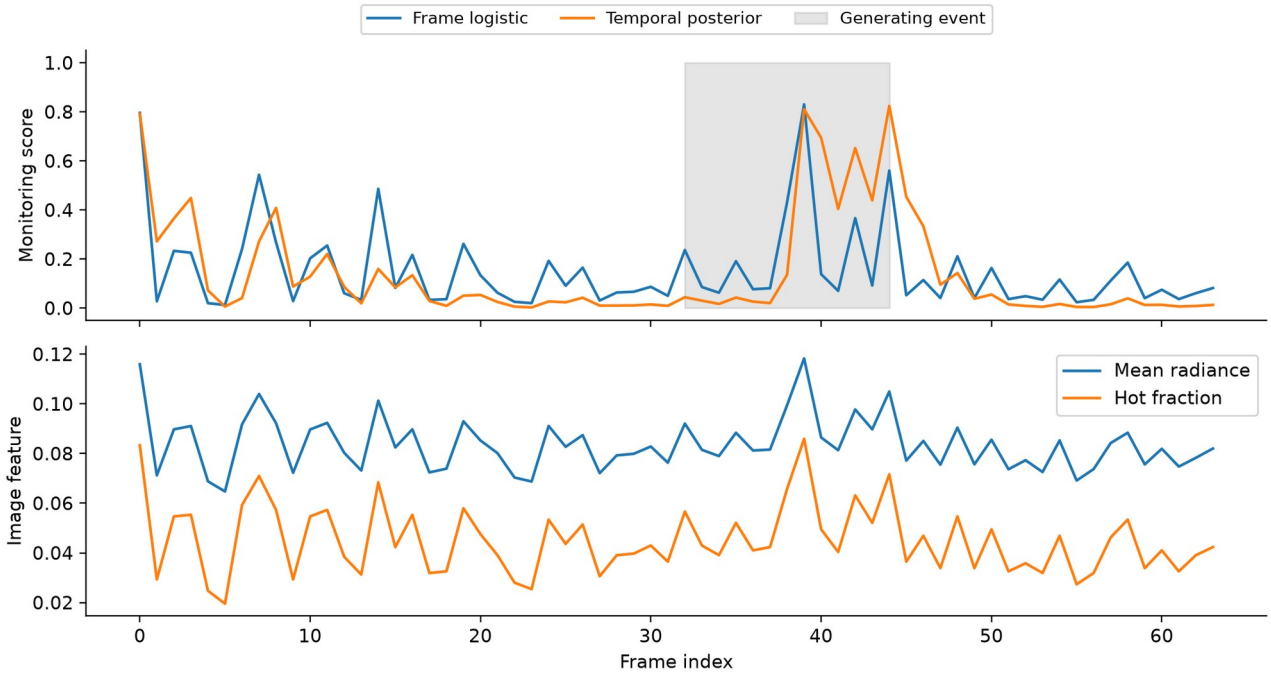


Figure 3. Executed causal monitoring trace for the same synthetic track. The generating interval is shaded; frame logistic scores and filtered probabilities use no future observations. Lower panel shows two directly calculated image features.

5 Thresholds and the monitoring estimands

The three monitored scores are mean radiance, learned frame probability and filtered state probability. Each receives its own threshold selected from ordinary validation frames at the ninety seventh percentile. This rule targets approximately three percent validation frame false positives without using test labels to optimize F1. The thresholds are approximately 0.10114, 0.43537 and 0.49140, respectively. Different score distributions require different numerical thresholds. Comparing all methods at a common value such as one half would confound temporal inference with an arbitrary score scale choice.

Frame precision is the fraction of alarmed frames whose generating state is changed, while frame recall is the fraction of changed frames alarmed. F1 combines those quantities. False positive rate divides ordinary alarmed frames by all ordinary frames. These are descriptive operating characteristics for the specified test population and event prevalence. They are not probabilities that a manufactured part is defective. A process event can span several frames, and many frame alarms can correspond to a single underlying episode, so frame and event level counts must be reported separately.

An event is detected if any alarm occurs during its generating interval. Delay is the number of frames between the generating start and the first in interval alarm, calculated only for detected events. This conditioning makes delay incomplete by itself: a method that misses difficult events can have an apparently favorable average among the easy ones it does detect. We therefore present event detections and delay together. Alarms occurring before or after an event do not count as detecting it, even if they are close in time. No retrospective tolerance window is added after viewing results.

Table 2 reports frame outcomes for all seventy test tracks. Table 3 gives event outcomes. There are sixty three generated events and 653 changed frames in the test set, leaving 3,827 ordinary frames. These counts permit direct reconstruction of precision and recall from the ledger. The validation and test frequencies differ

slightly because tracks and event durations are randomly drawn. Threshold transfer therefore has ordinary finite sample variability. We do not claim that a validation quantile guarantees the exact same false positive rate in every independently generated collection of tracks.

Table 2. Executed test frame outcomes over 4,480 frames with 653 changed and 3,827 ordinary states. Percentages use the stated denominators.

Score	Precision percent	Recall percent	F1	False frames
Brightness	55.94	24.50	0.3408	126
Frame logistic	55.79	24.35	0.3390	126
Temporal posterior	77.14	49.62	0.6039	96

Table 3. Event sensitivity and onset delay. Delay is conditional on detection; all methods evaluate the same sixty three generated events.

Score	Detected events	Missed events	Mean delay frames
Brightness	60	3	2.467
Frame logistic	60	3	2.417
Temporal posterior	57	6	3.649

Process monitoring studies have connected melt pool data to lack of fusion porosity [21], and directed energy deposition studies use learned segmentation and statistical analysis to infer process state [22]. Those are related but more demanding target definitions. The present labels are generated directly from the latent image event. There is no computed tomography, density measurement, destructive sectioning, defect annotation or part acceptance decision. We report no porosity accuracy and no manufacturing yield improvement. That boundary is essential because otherwise a strong frame score could be mistaken for validation of a downstream quality decision.

6 Executed results show both a gain and a cost

Brightness and frame logistic classification are nearly tied at the selected operating point. Brightness has precision 0.55944, recall 0.24502 and F1 0.34079. Logistic classification has precision 0.55789, recall 0.24349 and F1 0.33902. Both produce 126 false positive frames and a false positive rate of 0.03292. Thus this experiment does not support the claim that fitting five image measurements necessarily improves a carefully thresholded brightness baseline. The learned model is useful here as a probabilistic observation mechanism, but its independent frame ranking is not a clear performance advance.

Temporal filtering reaches precision 0.77143, recall 0.49617 and F1 0.60391. False positive frames fall to 96, with a false positive rate of 0.02508. The gain combines more true changed frames with fewer ordinary alarms. Figure 4 makes that joint change visible. It should not be translated into a general percentage improvement in manufacturing quality, because the metric concerns synthetic frame states. The image generator, event prevalence, validation threshold rule and fixed transition model all help determine the measured numbers.

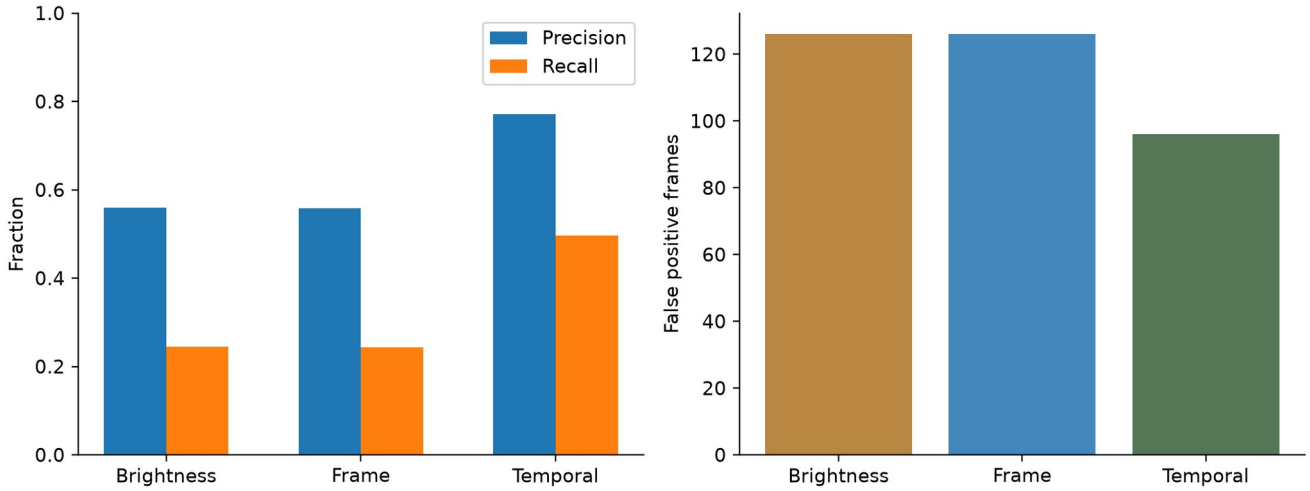


Figure 4. Executed test frame precision, recall and false alarm counts for brightness, learned frame scoring and temporal filtering. Thresholds are set independently at the ninety seventh percentile of ordinary validation scores.

The event results show why a single F1 value is insufficient. Brightness and logistic scoring each detect sixty of sixty three events, whereas the temporal filter detects fifty seven. Mean delay among detected episodes is 2.47 frames for brightness, 2.42 for logistic scoring and 3.65 for temporal filtering. Accumulating evidence suppresses isolated high scores but can fail to build enough probability during a weak or short event. The filter therefore produces purer sustained alarms while sacrificing some event sensitivity and speed. A monitoring objective that penalizes missed short events heavily might prefer a different threshold or transition model.

The frame counts expose the gain more concretely than a relative percentage. Brightness identifies 160 of the 653 changed frames; logistic scoring identifies 159. The temporal filter identifies 324. Thus filtering adds 165 true changed frame alarms relative to the learned frame rule while removing thirty ordinary alarms. These are counts from the same test tracks, not independent replicate means. The improvement is substantial for the specified frame objective, yet the filter still misses 329 changed frames. Reporting those residual misses is essential because an F1 near 0.60 can sound more complete than the underlying event coverage actually is.

At the episode level, sixty detections mean three missed generated events for brightness and logistic scoring. Fifty seven temporal detections mean six misses. A plausible explanation is that several short or weak episodes contain one high frame score that crosses the frame threshold but insufficient consecutive evidence to cross the temporal threshold. The saved traces allow that explanation to be inspected episode by episode. We do not attribute every miss to exactly that mechanism without an explicit case analysis, and we do not infer that increasing persistence would necessarily resolve it. More persistence can suppress onset as well as suppress noise.

The selected threshold convention also creates a useful qualification. Thresholds are matched by ordinary validation frame quantiles, not by identical test false positive counts. Test noise and track composition lead to a lower temporal false positive rate than the frame methods. Part of the final performance difference therefore reflects the full validation to test decision procedure, including score distributions and finite quantile estimates. It is not a comparison at a mathematically identical population false positive rate. That is an appropriate operational experiment, but it differs from interpolating complete test curves to a common error rate after outcomes are known.

7 A temporal order control isolates the mechanism

To test whether the improvement depends on event order, the program independently permutes the sixty four frame positions within each test track. Frame probabilities and true state labels are permuted together, preserving their instantaneous association and total counts. The Bayesian filter is then applied to the permuted score sequence using the original temporal threshold. This is an executed negative control, not a hypothetical proposed experiment. It removes sustained runs while retaining exactly the same set of image derived classifier scores in every track.

Temporal F1 falls from 0.60391 in the original order to 0.21140 after permutation. Precision becomes 0.47090 and recall 0.13629. The shuffled false positive rate is 0.02613, close to the ordered value, so the dominant loss is the failure to accumulate sufficient evidence on changed frames that no longer arrive in sustained episodes. Figure 5 pairs feature overlap with this temporal order comparison. The result supports the interpretation that temporal organization, rather than merely an alternate monotone transformation of the frame score, supplies the additional information.

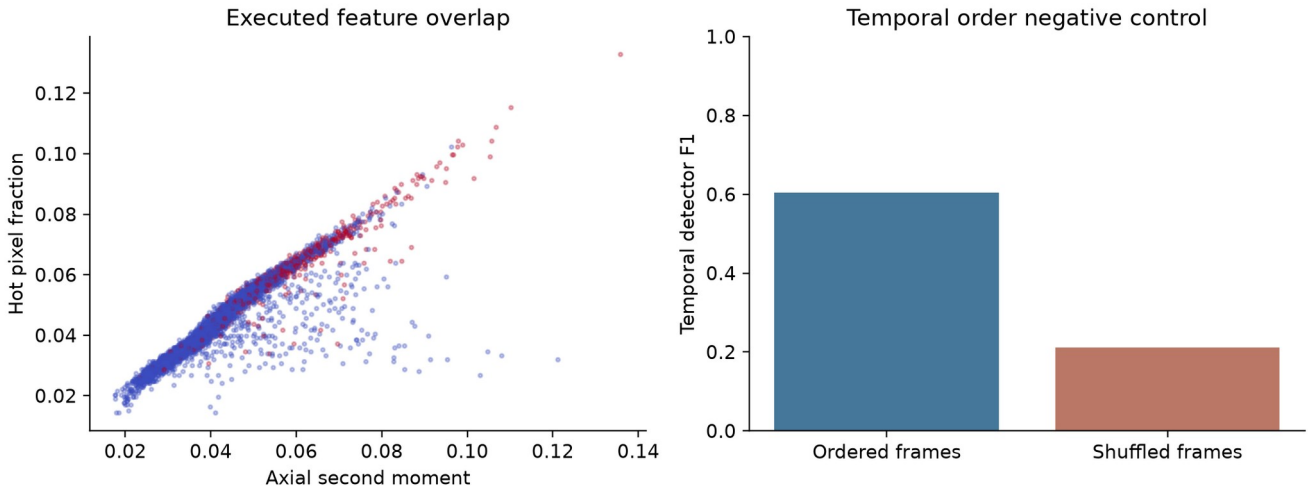


Figure 5. Executed image feature overlap and temporal order negative control. Test scores and states are permuted together within tracks; reduced filtered F1 identifies dependence on sustained ordering, without giving shuffled event delay a physical interpretation.

The shuffled control does not create a physically plausible alternate scan. A real process cannot generally reorder its frames without changing material evolution. Its value is analytical: it preserves the frame level evidence while disrupting the assumption used by the filter. We report only frame metrics for this control. Event delay under a scrambled sequence would not retain the meaning of elapsed time from a contiguous process onset. The script stores additional diagnostic fields, but they should not be interpreted as a second physically meaningful event detection experiment.

The control also does not prove that the chosen transition probabilities are optimal. Many persistence models could exploit runs of evidence, including run length rules, semi Markov durations or recurrent predictors. An event duration model matched exactly to the generator might improve sensitivity or delay, but testing it would require an additional implementation and independent comparison. The present filter deliberately uses a simple approximate two state mechanism. Its success on ordered frames and deterioration after permutation establish usefulness of persistence under these assumptions, without claiming the best possible sequential detector.

A second negative finding concerns maximum radiance. Its learned coefficient is negative after accounting for the other standardized features. This is consistent with occasional off center flashes being less specific than broad hot area, but coefficient signs in correlated regression do not establish a physical cause. Removing peak radiance or spatter would require additional fitted controls. We therefore treat the sign as an inspectable model property and a plausible interpretation of the generator, not as an executed ablation or a laboratory finding about the best camera statistic.

8 Relation to spatial representation and manufacturing physics

MeltpoolEmbedder develops orientation invariant information rich representations [4], while adaptable segmentation can reduce the annotation burden across different additive processes [5]. A laterally moving pool can rotate or change perspective, so invariance is a meaningful spatial issue. Our image axes remain aligned with the simulated scan and the pool center has a fixed nominal position. Temporal filtering does not solve rotation, magnification change or camera perspective. Its contribution would have to be combined with a suitable spatial representation before use in an unconstrained camera installation.

Engineering guided learning emphasizes features that reflect process variation and neighborhood relationships [6]. The present experiment is compatible with that principle but much simpler. It uses no three dimensional build neighborhood, layer history or machine geometry. A track contains a short numerical event in one optical stream. The gain from accumulation therefore should not be attributed to learning the spatiotemporal relationship between melt signatures and internal density. It arises from a predefined event persistence model applied to a fitted frame score, with truth supplied by the generator.

Multi source and multimodal workflows face several additional alignment problems. Different sensors may observe different physical volumes, camera exposure can integrate over changing scan speed, and post build quality measurements may correspond to regions that span several process frames. A visual monitoring system for melt pool detection demonstrates how spatial extraction can be embedded in actual acquisition [23]. Our coordinates and timestamps are exact by construction. This grants the filter an advantage unavailable automatically in practice: every frame belongs unambiguously to one track and one hidden interval.

The thermal interpretation is equally limited. Pixel radiance is not temperature without emissivity, spectral response and exposure calibration. A saturated bright point can behave differently from the unsaturated Gaussian fields used here. The synthetic camera has independent additive noise, no blur and no rolling shutter. The generated spatter flash is narrower than the central pool but has no ballistic trajectory or occlusion. These omissions simplify the scientific question and should guide what would need to be added before treating this model as a basis for hardware design.

Most importantly, sustained morphology is not synonymous with damage. A planned change in scan strategy, a corner turn or an intentional power adjustment can also persist. Conversely, a damaging pore formation event can be short. The state labels are chosen to study temporal evidence under weak frame separation, not to encode the true causal path from process conditions to defects. Any translation to part quality would require independent quality labels, a traceable spatial mapping and a decision analysis that distinguishes benign planned transitions from adverse events of interest.

There is one additional spatial simplification worth separating from temporal performance. The synthetic pool remains visible throughout every track. We do not generate smoke obscuration, recoater occlusion, lens contamination or complete saturation. A missing or unreadable frame would require an observation rule, such as skipping the likelihood update while retaining the transition prediction. Assigning it a low event

probability could instead create false reassurance. Because no missing frames occur here, neither treatment is evaluated, and the temporal improvement should not be interpreted as robustness to sensor dropout.

The event generator also holds duration independent of feature strength. In practice, a rapidly developing instability may change both how long a signal persists and how distinctive its individual frames become. Such dependence would alter the delay tradeoff. For example, long weak events and short strong events could favor different filters even at the same average event prevalence. The transparent generator makes this limitation visible and offers a direct route for later experiments, but those potential variants remain outside the executed evidence. The present conclusion concerns the exact joint duration and appearance model specified above.

9 Operational choices implied by the delay tradeoff

A monitoring system needs a time scale for action. If a controller can tolerate several frame delays while avoiding unnecessary interventions, the higher precision of temporal filtering may be useful. If the monitored event develops faster than the accumulation window, the delay may be unacceptable. Our result provides no physical time conversion, so it cannot choose between those applications. It instead demonstrates why reporting only a frame confusion matrix is inadequate whenever the intended use includes an intervention whose value depends on how early the alarm arrives.

The appropriate threshold also depends on event prevalence and consequences. We choose a validation false positive quantile to create a consistent comparison, not because three percent is an industrial acceptance limit. A lower threshold would generally increase recall and false alarms, while a higher threshold would reduce both. The relevant tradeoff may involve operator workload, automatic process interruption or retrospective quality review. Each requires a different cost definition. The fixed threshold experiment supplies one operating point and should not be described as a complete receiver operating characteristic study.

Tracks without events are particularly important in operational monitoring because a detector may otherwise learn that every sequence must contain an alarm. The generator includes such tracks explicitly. Nevertheless, the test set has a much higher event frequency than some high quality production environments might have. Precision would change under a different prevalence even if conditional sensitivity and false positive rate stayed the same. We therefore report the exact track and frame counts. We do not extrapolate the observed precision to a factory where adverse events may be rare and nonstationary.

The state filter can also continue to assign high probability after the generating event ends. Such persistence can be desirable for a human review alarm but costly for closed loop control. Our frame metrics count these trailing alarms as ordinary frame false positives, which ensures that temporal smoothness is not rewarded automatically. A part level alarm policy with hysteresis, reset rules or minimum duration could alter that accounting. It would need its own explicit definition and evaluation; none is added implicitly to improve the appearance of the reported temporal trace.

The experiment suggests a useful division of labor in monitoring design. Spatial learning should be evaluated for its ability to extract informative frame evidence, while temporal logic should be evaluated for its effects on delay, episode sensitivity and nuisance persistence. A joint recurrent model may ultimately perform better, but a modular baseline reveals which information source supplies an improvement. Here the spatial learner offers no improvement over brightness, while temporal organization produces a substantial frame level gain and a measurable event level cost. That contrast is the central scientific finding.

10 Reproducible evidence and conclusions

The source script generates all training, validation and test images from recorded seeds, fits the logistic coefficients, sets thresholds using validation negatives and writes a complete monitoring ledger. A second CSV preserves test feature values. The JSON file records coefficient values, standardization constants, transition choices, thresholds and summary metrics. NumPy supplies array operations [24], SciPy supplies numerical optimization [25], and Matplotlib renders the evidence figures [26]. No external dataset, proprietary camera record or unarchived pretrained model is required to reproduce the numerical result.

The complete procedure uses 9,600 frames across 150 tracks. Only the training split estimates spatial coefficients, only the validation split chooses operating thresholds, and only the test split determines the reported performance. The time permutation control uses the already generated test scores and a recorded pseudorandom permutation. This design separates learning, threshold selection and evaluation while retaining track level independence. It does not estimate variability across many independently generated datasets; the reported metrics describe one reproducible split and should be read with that finite sample scope.

Because adjacent test frames share event membership, uncertainty cannot be assessed as though all 4,480 frame decisions were independent Bernoulli trials. A frame bootstrap would break that dependence and could understate variability. Track based resampling would be more appropriate for uncertainty about this generator's collection of tracks, but it would still not cover uncertainty about material physics or camera installation. We report exact counts and descriptive metrics without a misleadingly precise significance claim. The independent unit in the design is the generated track, and the manuscript preserves that unit in every split.

The image archive itself is intentionally limited to the rendered numerical examples and the reproducible generator. Storing every synthetic pixel is unnecessary because the seed and exact code recreate the arrays, while the compact ledgers preserve the decisions needed to verify summary metrics. An independent reproduction should first compare the saved feature values, then classifier scores and thresholds, and finally alarm counts. Small floating point differences near a threshold can affect an individual frame, so exact coefficient and threshold values are kept in JSON rather than relying on rounded text in a table.

Finally, the method's modest size makes its failure modes interpretable. The classifier can fail because the spatial features overlap, the temporal filter can fail because evidence arrives too briefly, and the operating rule can fail because its selected threshold is too conservative for the intended cost. These are separate design questions. A larger recurrent neural network could merge them into one fitted representation, but that would not eliminate the need to define event onset, permissible delay and false alarm burden. This numerical study gives those definitions priority so that an apparent improvement has a clear operational meaning.

Table 4. Sequential model constants, controls and evidence limits.

Item	Specification	Meaning
Transition probabilities	$0 \rightarrow 1$: 0.018; $1 \rightarrow 1$: 0.88	Fixed approximate persistence model
Training prior	0.1515625	Removed from score odds before update
Thresholds	0.10114; 0.43537; 0.49140	Brightness; frame; temporal
Order control	Filtered F1 0.2114	Executed within track permutation

Item	Specification	Meaning
Evidence boundary	Generated morphology events	No porosity or hardware validation

A practical computational advantage of the two state update is its small memory requirement. Once the frame score is available, filtering needs only one previous posterior probability and fixed transition constants. The image feature extraction and frame model dominate the arithmetic for this small implementation. We make no latency claim because no deployment hardware or acquisition pipeline is timed. Algorithmic simplicity is a property of the implemented calculation, whereas real time operation would require end to end measurements including image transfer, synchronization and intervention interfaces.

The study concludes that persistence can be valuable evidence even when frame classification is only moderately informative. It also shows that better frame F1 can coexist with slower onset recognition and fewer detected events. The shuffled order control links the gain to temporal organization, while the brightness baseline prevents a weak spatial learning gain from being overstated. These results support an explicit delay aware evaluation of monitoring systems. They remain evidence about generated morphology states, and any claim about manufacturing defects would require physical measurements and independently established quality targets.

References

[1] Ross Cunningham, Cang Zhao, Niranjan Parab, Christopher Kantzos, Joseph Pauza, Kamel Fezzaa, Tao Sun, Anthony D. Rollett. Keyhole threshold and morphology in laser melting revealed by ultrahigh-speed x-ray imaging. Science, 363, 849-852, 2019. doi:10.1126/science.aav4687.

[2] Aiden A. Martin, Nicholas P. Calta, Saad A. Khairallah, Jenny Wang, Phillip J. Depond, Anthony Y. Fong, Vivek Thampy, Gabe M. Guss, et al.. Dynamics of pore formation during laser powder bed fusion additive manufacturing. Nature Communications, 10, 1987, 2019. doi:10.1038/s41467-019-10009-2.

[3] Cang Zhao, Kamel Fezzaa, Ross W. Cunningham, Haidan Wen, Francesco De Carlo, Lianyi Chen, Anthony D. Rollett, Tao Sun. Real-time monitoring of laser powder bed fusion process using high-speed X-ray imaging and diffraction. Scientific Reports, 7, 3602, 2017. doi:10.1038/s41598-017-03761-2.

[4] Zijue Chen, Heng Wang, Dayalan Gunasegaram. MeltpoolEmbedder: An information-rich, orientation-invariant deep learning framework for efficient melt pool feature extraction in metal additive manufacturing. Additive Manufacturing, 109, 104900, 2025. doi:10.1016/j.addma.2025.104900.

[5] Jiahong Chen, Hui Li, Shengnan Shen, Yingjie Wang, Rongzhe Ma, Haoyuan Chen, Hong Yin, Liwei Xia. Universal light SAM algorithm for in situ melting pool monitoring of L-DED/L-PBF/PAM processes. Journal of Computational Design and Engineering, 12, 35-48, 2025. doi:10.1093/jcde/qwaf057.

[6] Siqi Zhang, Hui Yang, Zhuo Yang, Yan Lu. Engineering-Guided Deep Learning of Melt-Pool Dynamics for Additive Manufacturing Quality Monitoring. Journal of Computing and Information Science in Engineering, 24, 101002, 2024. doi:10.1115/1.4066026.

[7] Wayne E. King, Holly D. Barth, Victor M. Castillo, Gilbert F. Gallegos, John W. Gibbs, Douglas E. Hahn, Chandrika Kamath, Alexander M. Rubenchik. Observation of keyhole-mode laser melting in laser powder-bed fusion additive manufacturing. Journal of Materials Processing Technology, 214, 2915-2925, 2014. doi:10.1016/j.jmatprotec.2014.06.005.

[8] Saad A. Khairallah, Andrew T. Anderson, Alexander Rubenchik, Wayne E. King. Laser powder-bed fusion additive manufacturing: Physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones. Acta Materialia, 108, 36-45, 2016. doi:10.1016/j.actamat.2016.02.014.

[9] Manyalibo J. Matthews, Gabe Guss, Saad A. Khairallah, Alexander M. Rubenchik, Philip J. Depond, Wayne E. King. Denudation of metal powder layers in laser powder bed fusion processes. Acta Materialia, 114, 33-42, 2016. doi:10.1016/j.actamat.2016.05.017.

[10] Nobuyuki Otsu. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9, 62-66, 1979. doi:10.1109/tsmc.1979.4310076.

[11] John Canny. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8, 679-698, 1986. doi:10.1109/tpami.1986.4767851.

[12] Arthur E. Hoerl, Robert W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12, 55-67, 1970. doi:10.1080/00401706.1970.10488634.

[13] Leo Breiman. Random Forests. Machine Learning, 45, 5-32, 2001. doi:10.1023/a:1010933404324.

[14] Luke Scime, Jack Beuth. Anomaly detection and classification in a laser powder bed additive manufacturing process using a trained computer vision algorithm. Additive Manufacturing, 19, 114-126, 2018. doi:10.1016/j.addma.2017.11.009.

[15] Kristen J. Hernandez, Thomas G. Ciardi, Rachel Yamamoto, Mingjian Lu, Arafath Nihar, Jayvic Cristian Jimenez, Pawan K. Tripathi, Brian Giera, et al.. L-PBF High-Throughput Data Pipeline Approach for Multi-modal Integration. Integrating Materials and Manufacturing Innovation, 13, 758-772, 2024. doi:10.1007/s40192-024-00368-0.

[16] Zhiyong Zou, Kai Zhang, Tingting Liu, Jiansen Li, Zhiguang Zhu, Huiliang Wei, Yuxian Lu, Wenhe Liao. In situ monitoring with melt pool data based on multi-signal fusion method in laser powder bed fusion. Measurement, 234, 114877, 2024. doi:10.1016/j.measurement.2024.114877.

[17] A. Wald. Sequential Tests of Statistical Hypotheses. The Annals of Mathematical Statistics, 16, 117-186, 1945. doi:10.1214/aoms/1177731118.

[18] E. S. PAGE. CONTINUOUS INSPECTION SCHEMES. Biometrika, 41, 100-115, 1954. doi:10.1093/biomet/41.1-2.100.

[19] Leonard E. Baum, Ted Petrie, George Soules, Norman Weiss. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. The Annals of Mathematical Statistics, 41, 164-171, 1970. doi:10.1214/aoms/1177697196.

[20] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering, 82, 35-45, 1960. doi:10.1115/1.3662552.

[21] Sam Coeck, Manisha Bisht, Jan Plas, Frederik Verbist. Prediction of lack of fusion porosity in selective laser melting based on melt pool monitoring data. Additive Manufacturing, 25, 347-356, 2019. doi:10.1016/j.addma.2018.11.015.

[22] Reza Asadi, Antoine Queguineur, Olli Wiikinkoski, Hossein Mokhtarian, Tommi Aihkisalo, Alejandro Revuelta, Iñigo Flores Ituarte. Process monitoring by deep neural networks in directed energy deposition: CNN-based detection, segmentation, and statistical analysis of melt pools. Robotics and Computer-Integrated Manufacturing, 87, 102710, 2024. doi:10.1016/j.rcim.2023.102710.

[23] Trong-Nhan Le, Min-Hsun Lee, Ze-Hong Lin, Hong-Chuong Tran, Yu-Lung Lo. Vision-based in-situ monitoring system for melt-pool detection in laser powder bed fusion process. Journal of Manufacturing Processes, 68, 1735-1745, 2021. doi:10.1016/j.jmapro.2021.07.007.

[24] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, et al.. Array programming with NumPy. Nature, 585, 357-362, 2020. doi:10.1038/s41586-020-2649-2.

[25] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, et al.. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17, 261-272, 2020. doi:10.1038/s41592-019-0686-2.

[26] John D. Hunter. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9, 90-95, 2007. doi:10.1109/mcse.2007.55.