FLOWCTRL: ROBUST TERMINAL CONTROL OF FLOW MODELS FOR ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Detecting defects by restoring the test image requires keeping every normal pixel and none of the defect, while which pixels are defective is unknown. We cast this as one control problem on a flow matching model trained on normal images, and propose FlowCTRL: the flow generates a normal image from a Gaussian draw, a control steers it towards the observation, and a per-pixel belief sets how far the observation is trusted. Because steering pulls an unflagged defect into the restoration, the method measures the belief first, on probe paths that no control touches, and then steers the flow with a closed-form controller that blends the model's endpoint prediction with the observation pixel by pixel. The restoration keeps the observation where it is trusted and the model's own completion elsewhere, and its residual is the anomaly map. FlowCTRL uses only forward passes of a pixel-space flow model, without pretrained features, at 0.28 s per image. On MVTec-AD and VisA, it shows remarkable pixel-level localization performance matching or surpassing methods built on pretrained feature extractors including DINOv2, and presents strong image-level anomaly detection ability.

1 INTRODUCTION

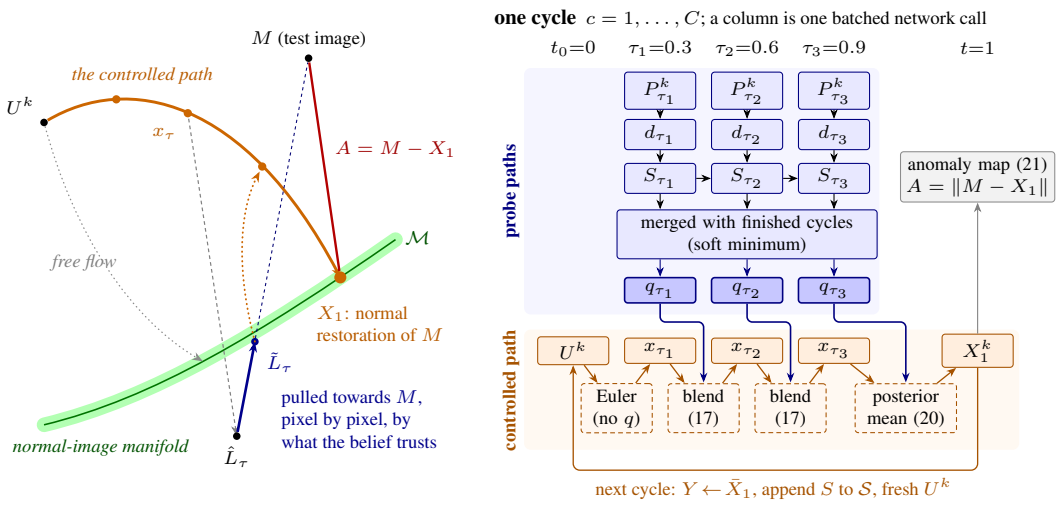
Industrial inspection has to find defects that are rare, diverse and often small, from training data that contain only normal images (Bergmann et al., 2019). A direct way is to restore the test image, producing the normal image it should have been, and to report where the two differ. The restoration must reproduce every normal pixel, or the difference flags normal detail, and must not reproduce the defect, or the difference misses it; yet which pixels are defective is exactly what is unknown.

Existing methods and limitations. Reconstruction networks settle this through what they are able to reproduce, and a network that restores normal detail well tends to restore the anomaly too, the "identical shortcut" of You et al. (2022). A generative model trained on normal images describes normality explicitly, but its sample is *a* normal image, not the normal version of the observation, and steering the sample towards the observation everywhere copies the defect as well. Most current methods compare the features of a large pretrained network rather than pixels (Roth et al., 2022; Guo et al., 2025).

Anomaly detection as control. We make this tension explicit (Figure 1). A flow matching model (Lipman et al., 2023; Liu et al., 2023a) trained on normal images generates a normal image from a Gaussian draw, a control steers the generation towards the observation at a cost, and a per-pixel *belief* that the pixel is anomalous sets how much of the observation the control may use. Modelling the observation as a normal image plus a sparse, contiguous anomaly plus noise makes this one optimization problem over the control and the belief (Section 3.1). Its solution is a restoration that equals the observation where a pixel is trusted and the model's own completion where it is not; its residual is the anomaly map.

Measure first, then steer. The two unknowns cannot be alternated: a defect that the belief has not yet flagged is trusted, pulled towards the observation and copied into the restoration, and its evidence vanishes (Lemma 1). The product rule gives an order instead. FlowCTRL first measures the belief on *probe paths* from fresh Gaussian draws to the observation, which no control touches; in the first cycle, their energy is the velocity mismatch of Chen et al. (2026) (Proposition 4). It then steers the

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.



(a) what the controller does

(b) one cycle of the run

Figure 1: **Anomaly detection as a control problem**. **(a)** A random draw with the flow model (grey) lands at an ordinary normal image, not at the normal version of M. We steer it instead. At a state x_τ the controller reads the model's own endpoint estimate $\hat{L}_\tau$, pulls it towards M by as much as each pixel can be trusted (belief), and recomposes the next state (dotted). The run ends at X_1, the normal restoration of M, and $A = M - X_1$ is the anomaly map. **(b)** One cycle unrolled; a column is one batched call on $K_c{+}K_p$ states. Belief q on anomaly is not read off the controlled path but measured on *probe* paths, which no control touches. Their residuals against M are pooled, filtered, merged with the finished cycles and turned into a belief q_{τ_j}, refresh at every probe time, and crossing to the controller. A run repeats the cycle C times. Figure 4 draws detailed and explicit steps of (a).

flow with that belief by a closed-form, per-pixel blend of the model's endpoint prediction and the observation at each probe time, and restoration cycles re-aim the probes at the previous restoration. A run takes 8 batched network calls and no back-propagation.

Contributions. (1) *Formulation.* We cast anomaly localization and restoration as one control problem on a flow model trained on normal images, relax the anomaly support to a belief whose threshold at $\frac{1}{2}$ is exact for a fixed restoration (Proposition 1), and show why the belief must be measured before the control acts. (2) *Method.* FlowCTRL pairs an observer on uncontrolled probe paths with a closed-form controller, at 0.28 s per image. (3) *Empirics.* In pixel space at 128×128 and without pretrained features, FlowCTRL's pixel AP, F_1-max and AUROC on MVTec-AD are exceeded only by two methods built on DINOv2 features. Removing the control, or measuring the belief on the controlled path instead of the probe paths, lowers pixel AP by more than half on both datasets (Table 2).

2 Related Works

Anomaly detection. Embedding methods model pretrained features of normal images as Gaussian patch statistics (Defard et al., 2021), a memory bank (Roth et al., 2022) or a normalizing flow (Gudovskiy et al., 2022); distillation methods compare a student with its pretrained teacher (Deng & Li, 2022; Tien et al., 2023). Reconstruction methods restore the image or its features and read the residual (You et al., 2022; Zhang et al., 2025; Guo et al., 2025; Luo et al., 2025), some trained to remove synthetic anomalies (Zavrtanik et al., 2021; Zhang et al., 2023) and some with a diffusion model as the reconstructor (He et al., 2024b; Feng et al., 2025). Flow matching has entered anomaly detection through modified training (Li et al., 2025a; Beizaee et al., 2025; Li et al., 2025b) and through the velocity field of a normally trained model along paths towards the test image (Chen et al., 2026). FlowCTRL keeps such paths as its observer, but uses their evidence as the belief of a control problem whose solution is a restoration.

Control in flow matching. A flow model generates by integrating an ODE, which can be steered through its source, its velocity or its weights. D-Flow (Ben-Hamu et al., 2024) and FlowGrad (Liu et al., 2023b) differentiate a cost on the sample through the solver to optimize the source or the velocity; training-free inverse-problem solvers correct each step towards the measurement (Pokle et al., 2024; Martin et al., 2025); RF-Inversion (Rout et al., 2025) steers with a linear quadratic regulator; Adjoint Matching (Domingo-Enrich et al., 2025) fine-tunes the weights by stochastic optimal control; and inpainting keeps the observed pixels under a known mask (Lugmayr et al., 2022). All of them are given what to steer towards; in anomaly detection, which pixels to trust is itself unknown. FlowCTRL infers it as a belief from uncontrolled paths and weighs the observation pixel by pixel in closed form, without back-propagation through the solver.

3 METHOD

The task. We are given one velocity field $v_\theta(x, t, c)$ trained by flow matching (Lipman et al., 2023; Liu et al., 2023a) on the normal images of all categories, with the category label c as conditioning. Training uses the conditional optimal-transport path $x_t = (1-t)\,U + t\,X$, $U \sim \mathcal{N}(0, I)$, $t \in [0, 1]$, with target velocity $X - U$, so that $t = 0$ is noise and $t = 1$ is data. The label is fixed throughout a run and we drop it from the notation. The model is never updated after training, and no anomalous image is ever seen. At test time we observe a test image M of resolution $[H, W]$ and model it as

$$M_i \;=\; L_i + A_i + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, s^2 I_C), \qquad A_i \in \mathbb{R}^C, \tag{1}$$

where i is pixel index, L is an unknown normal image, A is an anomaly (heatmap) that is large where M is anomalous, supported on an unknown and typically small set of pixels, and ε is noise. We want a restoration $X_1 \approx L$ and a per-pixel map of where M is not explained by it. Both come out of one problem, which we set up first.

3.1 ONE CONTROL PROBLEM FOR THE ANOMALY MAP AND THE RESTORATION

The general optimization. Equation (1) says three things about the observation: where there is no anomaly, M is the normal image up to sensor noise. The anomaly occupies few pixels, and those pixels form contiguous regions. It says one thing about L that it is a *normal* image, and the only description of normal images we have is the flow model. We therefore let the flow produce L from a Gaussian draw U, steered by a control effort a_t at a price.

$$\begin{aligned}\min_{a,\,A}\; \sum_i \Big[\underbrace{\frac{\|M_i - L_i - A_i\|^2}{2s^2}}_{\text{noise}} + \underbrace{\lambda\, \mathbf{1}\{A_i \neq 0\}}_{\text{sparsity}} \Big] + \underbrace{\mu \sum_{i \sim j} \big|\mathbf{1}\{A_i \neq 0\} - \mathbf{1}\{A_j \neq 0\}\big|}_{\text{contiguity}} \\ + \underbrace{\int_0^1 \tfrac{1}{2}\|a_t\|^2 \,\mathrm{d}t}_{\text{normality}}, \qquad \text{s.t. } \dot{x} = v_\theta(x, t) + a_t, \quad x_0 = U, \quad L = x_1.\end{aligned} \tag{2}$$

The contiguity term charges μ for every pair of 4-neighbours $i \sim j$ that the support separates. The normality term is the effort of steering the flow, which vanishes exactly when L is the flow's own sample from U and grows as L is pulled away from it, so it acts as a prior that prefers the images generated by the flow. The control effort a is necessary in (2). Without it, (2) is solved by $L = M$ and $A = 0$ at no cost, and nothing is detected. With it, L follows M only as far as the flow can be steered, and the further M departs from the images generated by the model, the more reproducing it costs. Where that is not worth the control effort, L keeps the model's own completion, the residual stays, and calling the pixel anomalous at the price λ becomes the cheaper explanation. This trade-off, between the cost of explaining a pixel as normal and the price of calling it anomalous, is what the rest of the method computes.

Relaxing the anomaly. The anomaly can be minimized out of (2) before the control, $\min_{a,A} = \min_a \min_A$. At any control a, and hence any L, the anomaly enters the noise term pixel by pixel, while the sparsity and contiguity terms see only its support $z_i = \mathbf{1}\{A_i \neq 0\}$. A pixel with $z_i = 0$ has $A_i = 0$ and pays the residual, and a pixel with $z_i = 1$ is best served by $A_i = M_i - L_i$, which

eliminates the quadratic, and pays λ. The control effort does not involve A and passes through unchanged, so (2) becomes a problem in the control and the binary support $z \in \{0,1\}^{HW}$ alone,

$$\min_{a,\,z} \int_0^1 \tfrac{1}{2}\|a_t\|^2 \,\mathrm{d}t + \sum_i \big[(1-z_i)\,d_i(L) + z_i\,\lambda\big] + \mu \sum_{i\sim j} |z_i - z_j|, \quad d_i(L) = \frac{\|M_i - L_i\|^2}{2s^2}, \tag{3}$$

attained at $A^\star = z \odot (M - L)$. $d_i(L)$ is the data residual term under the hypothesis that pixel i is normal, and λ is the price of calling it anomalous. Without the coupling ($\mu = 0$) each pixel chooses alone and pays the smaller of the two, equivalently a truncated quadratic:

$$\min_{z} \sum_i \big[(1-z_i)\,d_i + z_i\,\lambda\big] \;=\; \sum_i \min\{d_i(L), \lambda\}. \tag{4}$$

What we want instead is a *relaxed* decision, in which a pixel whose evidence is marginal is half-trusted. Relaxing z to the cube $[0,1]^{HW}$ changes nothing by itself, since the objective is linear in z up to the coupling, and the minimum is still attained at a vertex (Proposition 1). We write $q_i \in [0,1]$ for it and add an entropy at temperature T, which gives per pixel the free energy

$$\psi(d_i, q_i) = (1-q_i)\,d_i + q_i\lambda + T\big[q_i \log q_i + (1-q_i)\log(1-q_i)\big], \tag{5}$$

and with it the relaxed problem:

$$\min_{a,\; q\in[0,1]^{HW}} \int_0^1 \tfrac{1}{2}\|a_t\|^2 \,\mathrm{d}t + \sum_i \psi\big(d_i(L), q_i\big) + \mu \sum_{i\sim j} |q_i - q_j|, \qquad L = x_1 \text{ as in (2)}, \tag{6}$$

q is called the *belief*: q_i is the amount by which pixel i is believed anomalous, and $1 - q_i$ the amount by which it is trusted. Problem (6) has two unknowns: the belief q, which says how anomalous each pixel is, and the control a, which steers the flow and so decides the normal image L. It couples the belief and the control only through the term $\sum_i (1-q_i)\,d_i(L)$, i.e. the residual of a pixel counts only as much as the belief trusts that pixel. If L were known, what remains would be a problem in q alone (Section 3.2); if q were known, it would be a problem in a alone (Section 3.3).

Why the belief comes first. The obvious way to solve the two together is to alternate: compute the belief from the current restoration, steer with it, and repeat. However, this fails. The control pulls every trusted pixel towards M. A defect that the belief has not flagged yet is trusted too, so it is pulled towards M, and its residual shrinks, the belief trusts it even more, and the next step pulls it harder. In the end the defect is copied into the restoration and is never flagged (Appendix A.2.3). The belief must therefore be computed from something the control cannot change.

The total cost $\mathcal{E}$ of (3) can be viewed in a probability perspective, $p \propto e^{-\mathcal{E}}$, which gives the probability of a control steered normal image L and a set of anomalous pixels z given the observation M. Like any joint probability, it splits into two parts (the product rule, Section A.2.2),

$$p(z, L \mid M) \;=\; p(z \mid M)\; p(L \mid M, z).$$

The first part $p(z \mid M)$ says which pixels are anomalous without choosing a restoration, and by looking at the normal images produced by the flow on its own without control, and measures how well they explain each pixel of M on average. The second part $p(L \mid M, z)$ is the normal image once we know which pixels to trust, by steering the flow towards M on the trusted pixels and left free on the others. So we first measure the belief on the flow without control, on paths that the control never touches (Section 3.2), and then steer the flow with that belief (Section 3.3). Within a cycle, the belief never looks at the paths being steered; the two parts meet only when the controller reads the belief and, across cycles, when the probes are aimed at the previous restoration (Section 3.4).

3.2 The observer: measuring where the model refuses the observation

The belief. For a fixed data residual energy d, the belief block of (6) is

$$q^\star(d) = \arg\min_{q\in[0,1]^{HW}} \;\sum_i \psi(d_i, q_i) \;+\; \mu \sum_{i\sim j} |q_i - q_j|, \tag{7}$$

which is strictly convex, so $q^\star$ is unique. Without the coupling ($\mu = 0$) it decides each pixel alone, in closed form:

$$q_i^\star(d_i) = \sigma\Big(\frac{d_i - \lambda}{T}\Big), \qquad \Phi(d_i) = \min_{q_i} \psi = -T \log\big(e^{-d_i/T} + e^{-\lambda/T}\big), \tag{8}$$

where σ is the sigmoid function and Φ is a soft minimum of d_i and λ (Appendix A.3.1). With $\mu > 0$ a defect is decided as a region rather than alone, and (7) is solved in the dual and by iterations, see Section A.3.2 for more details. What is left is the data residual energy d.

Probe paths. As Section 3.1 ruled out, the belief has to be measured without control. We therefore introduce a parallel observer consisting of *probe paths*, which are built from their own sources and no control ever touches. A probe path is the interpolant with source U and a target image Y. In the very first cycle, $Y = M$ i.e. the observation image. There are K_p probe paths in total.

$$P_\tau^k \;=\; (1-\tau)\,U^k + \tau\,Y, \qquad k = 1, \dots, K_p. \tag{9}$$

Since v_θ approximates $\mathbb{E}[X - U \mid x_t, t]$, and the flow path gives $x_t + (1-t)(X-U) = X$, the quantity $\hat{L}_t(x) = x + (1-t)\,v_\theta(x,t)$ is the model's estimate of the *data* endpoint of the path through x at time t. Using the data term of (3), the probes measure the residual between the prediction $\hat{L}_\tau(P_\tau^k)$ and the observation, in units of the noise scale that a state at time τ still carries:

$$d_{\tau,i}^k \;=\; \frac{\big\|M_i - \hat{L}_\tau(P_\tau^k)_i\big\|^2}{2\,s_\tau^2}, \quad s_\tau = 1 - \tau, \quad d_{\tau,i} = -T \log \frac{1}{K_p} \sum_{k=1}^{K_p} e^{-d_{\tau,i}^k/T}, \tag{10}$$

where d_τ is the soft minimum of all paths. Notably, the hypothesis being tested is about the observation, so the residual is taken against M *always*, even in later cycles where the probe line ends at a different target Y (Section A.3.5). The scale $s_\tau = 1 - \tau$ makes the measurements taken at different times commensurate, so that a prediction made at $\tau = 0.3$ is allowed a larger absolute error than one made at $\tau = 0.9$. The K_p probes at one time τ differ only in their source draw, and give K_p energies for the same pixel. The source is a nuisance variable of the first factor of Section 3.1, and the soft minimum in (10) integrates it out of the belief: in (8) it gives $\operatorname{logit} q_i = -\lambda/T - \log \frac{1}{K_p} \sum_k e^{-d_{\tau,i}^k/T}$, a Monte Carlo estimate of $-\lambda/T - \log \mathbb{E}_U e^{-d_{\tau,i}(U)/T}$, which averages the likelihood of the normal hypothesis over sources. A pixel therefore counts as explained (normal) as soon as some plausible source reproduces it (Section A.3.4).

Accumulation over probe times. Evidence from the j probe times is combined by a first-order filter. With $\gamma_j = 1 - e^{-\kappa(\tau_j - \tau_{j-1})}$,

$$S_{\tau_j} \;\leftarrow\; (1-\gamma_j)\,S_{\tau_{j-1}} + \gamma_j\,d_{\tau_j}, \qquad S \text{ initialized at } d_{\tau_1}. \tag{11}$$

This is the zero-order-hold discretization of $\dot{S} = \kappa(d - S)$ and is therefore *exact* if d is piecewise constant between measurements. The accumulated evidence does not depend on how finely the probe times are spaced, only on where they are, so changing J changes what was measured and not how it was averaged. The rate $\kappa = 4$ gives the last probe time about $\gamma = 0.70$ of the weight at the default spacing $\tau \in \{0.3, 0.6, 0.9\}$. Then, belief q_{τ_j} is obtained by solving (7) at $d = S_{\tau_j}$ (merged with the finished cycles, Section 3.4), and handed to the controller.

3.3 The controller: steering towards data consistency

The control problem. For a fixed belief, (6) is a control problem, and what it asks of a run is clearest with the support itself in place of the belief. Suppose for a moment that the anomaly support $z = \mathbf{1}\{A \neq 0\}$ were known. On a normal pixel ($z_i = 0$) the observation M *is* the clean image up to sensor noise, so the restoration $X_1 = L$ should end there, and putting anything else in its place would discard detail the observation already carries. On an anomalous pixel ($z_i = 1$) the observation says nothing about L_i at all by (1), so nothing should push the restoration anywhere and the flow should be left to generate. That is (3) at a known z, where λ and the coupling are constants, yielding a terminal-cost control problem whose cost is the data term summed over the normal pixels *only*,

$$\min_a \int_0^1 \tfrac{1}{2}\|a_t\|^2\,\mathrm{d}t \;+\; (1-z) \odot \frac{\|M - X_1\|^2}{2s^2}, \qquad \dot{x} = v_\theta(x,t) + a_t, \quad x_0 = U, \tag{12}$$

where $X_1 = x_1$ is the restoration the control produces. Equation (12) can be solved in closed form, yielding: (details in Section A.4.2)

$$a_t \;=\; \Big(\frac{\partial X_1}{\partial x_t}\Big)^{\top} \frac{(1-z)\odot\big(M - X_1\big)}{s^2}. \tag{13}$$

It is also the stationarity condition of (6) with the belief optimal rather than fixed, with $1-q^\star\big(d(X_1)\big)$ in place of $1-z$, since the minimum of (7) has derivative $1-q_i^\star$ in d_i (Appendix A.3.1), meaning that the belief enters the control only as the weight $1-q$ on a pixel's residual. Equation (13) is implicit, since X_1 is the state the control produces, and for a general velocity field evaluating it means an adjoint pass along the whole trajectory, which is time consuming.

Endpoint prediction and single action control. To avoid a whole trajectory tracing (Section A.4.3), we only focus on a family of single-step control problems indexed by the decision time $t \in \{\tau_1, \ldots, \tau_J\}$, i.e. the control-steered endpoint $\tilde{L}_{1,t}$ is estimated from the intermediate state x_t with the endpoint estimate $\hat{L}_t$ of Section 3.2:

$$\tilde{L}_{1,t} = x_t + s_t(v_\theta(x_t,t) + a) = \hat{L}_t(x_t) + s_t\, a, \quad s_t = 1-t. \tag{14}$$

A single action $s_t a$ holds over the interval, and costs $\frac{1}{2}\|a\|^2 s_t^2$ of effort. The quantity compared with M is not a finished image but the model's prediction of one. Therefore the data-consistency weight is annealed as s_t, reaching hard consistency exactly at $t=1$. In the control process, what we have instead of z is the time-dependent belief q_t of Section 3.2. Collectively:

$$\min_a \; \mathcal{J}(a) = \frac{s_t^2}{2}\,\|a\|^2 \;+\; (1-q_t)\,\frac{\|M-\tilde{L}_{1,t}\|^2}{2\,s_t^2}. \tag{15}$$

Setting the derivative of (15) to zero, i.e. $d\mathcal{J}/da = 0$, multiplying by s_t and collecting a gives $a\big[s_t^3 + s_t(1-q_t)\big] = (1-q_t)(M-\hat{L}_t)$, hence

$$a^\star \;=\; \frac{(1-q_t)\,\big(M-\hat{L}_t\big)}{s_t\,\big(s_t^2+1-q_t\big)}, \qquad \tilde{L}_{1,t} \;=\; \hat{L}_t + \frac{1-q_t}{s_t^2+1-q_t}\,\big(M-\hat{L}_t\big), \tag{16}$$

equivalently a per-pixel *blend* of the prediction and the observation,

$$\tilde{L}_{1,t} = \frac{\hat{L}_t + wM}{1+w}, \quad M - \tilde{L}_{1,t} = \frac{M-\hat{L}_t}{1+w}, \quad w = \frac{1-q_t}{s_t^2}. \tag{17}$$

Intuitively, when $q_t \to 0$, i.e. the pixel is trusted as normal, w gets large so that the blend approximates normal pixels on M. Instead, on anomalous pixels $w \to 0$, and it tends to simulate the flow model estimation $\hat{L}_t$. In this way, the control process is actually guiding the restoration towards an anomaly-free version of M while keeping its information.

Succeeding state. Recall that a state is the mixture of its two endpoint estimates,

$$x_t = (1-t)\,\hat{U}_t(x_t) + t\,\hat{L}_t(x_t), \quad \hat{L}_t(x) \;=\; x + (1-t)\,v_\theta(x,t), \qquad \hat{U}_t(x) = x - t\,v_\theta(x,t). \tag{18}$$

Moving from the current time τ_j to the next grid time τ_{j+1} produces the next state $x_{\tau_{j+1}}$. The source endpoint $\hat{U}_t$ enters the recomposition as the model currently estimates it, and is re-estimated from the new state at every call, and the controls drive it away from the draw the path started at.

A complete control process therefore goes like the following way. Starting from a Gaussian draw U and writing $t_0 = 0 < \tau_1 < \cdots < \tau_J < t_{J+1} = 1$,

$$x_0 = U, \qquad x_{\tau_{j+1}} \;=\; (1-\tau_{j+1})\,\hat{U}_{\tau_j}(x_{\tau_j}) \;+\; \tau_{j+1}\,\tilde{L}_{1,\tau_j}, \qquad j = 0, \ldots, J. \tag{19}$$

The first step carries no belief and is a plain Euler step. The belief q_{τ_j} used at step j is the one the observer has just refreshed at the same time (Section 3.2), before the blend (17) is formed. At the last step $t_{J+1} = 1$, the coefficient $(1-t_{J+1})$ vanishes, so (19) degenerates: the source endpoint drops out entirely and the state *is* the planned endpoint, $X_1 = \tilde{L}_{1,\tau_J}$. There is no further state to

build and no further measurement to wait for. However, the annealed weight of (17) reaches hard consistency only as $t \to 1$, while the last decision is taken at $\tau_J < 1$. The controller therefore takes that limit directly so that $w \to \infty$ to guarantee data consistency with M:

$$X_1 \;=\; (1 - q_{\tau_J}) \odot M \;+\; q_{\tau_J} \odot \hat{L}_{1,\tau_J}(x_{\tau_J}). \tag{20}$$

Thus output controlled restoration X_1 is exactly M where the pixel is trusted and exactly the model's own completion where it is not, with no step size and no scale left in it.

3.4 Restoration cycles

A run consists of C *cycles*, and a cycle runs K_p probe paths as the observer and K_c controlled paths as the controller. Both components are concatenated into one network call per time step. One cycle produces a restoration X_1. Repeating it, with the probe target Y moved to that restoration, gives a second measurement of the same support along paths that end somewhere the first cycle has already repaired. Cycle c probes towards $Y_c = \frac{1}{K_c}\sum_k X_1^{k,(c-1)}$ while still measuring every residual against M by (10). A pixel the first cycle managed to repair is then explained by the probe prediction and its evidence falls, and a pixel it failed to repair keeps its evidence. The cycle evidences $S^{(1)}, \dots, S^{(c-1)}, S^{(c)}$ are combined with soft minimum like in (10) (Section A.5.1). Each cycle draws its own sources by default and re-runs the controlled paths from scratch. The probe target is the only way the controller reaches the observer, so within a cycle the belief never depends on the controlled states, as Section 3.1 requires (Proposition 7).

Readout. A run ends with K_c restorations X_1^k. By (20), $M - X_1^k = q_{\tau_J} \odot \big(M - \hat{L}_{1,\tau_J}^k\big)$, where $\hat{L}_{1,\tau_J}^k$ is the endpoint prediction of path k at the last step, and the belief q_{τ_J} is the same for every path. Each residual is therefore the relaxed anomaly $q \odot (M - L)$ of Section 3.1, with the completion of path k in place of L. The anomaly map is the largest of them (Section A.5.3):

$$A_i = \sqrt{2\, e_i}, \qquad e_i = \max_{k=1,\dots,K_c} e_i^k, \qquad e_i^k = \tfrac{1}{2}\big\|M_i - X_{1,i}^k\big\|^2. \tag{21}$$

The maximum is a choice of the readout rather than a consequence of (6). Where the belief trusts a pixel ($q \approx 0$), every residual is close to zero and the maximum adds nothing. Where the belief rejects a pixel, the paths complete it differently, and the maximum keeps the completion that departs most from the observation, so a defect is not hidden by a path that happens to reproduce part of it. See Figure 1 for an intuitive overview of the method run. Post-processing is fixed to a 9×9 Gaussian blur with $\sigma = 3$ on A, and the image-level anomaly score is the mean of the top 1% of A.

4 Experiments

4.1 Setup

Data and model. We evaluate on **MVTec-AD** (Bergmann et al., 2019) (15 categories: 10 objects, 5 textures; 3,629 training and 1,725 test images) and **VisA** (Zou et al., 2022) (12 categories; 8,659 training and 2,162 test images), whose training sets contain only normal images. All images are uniformly resized to 128×128 and normalized according to normal images' mean and standard deviation, without additional augmentation. We report AUROC, AP and F_1-max at image and pixel level, and AUPRO up to a false-positive rate of 30% (Bergmann et al., 2020). For the flow matching model, we adopt a U-Net with attention (Lipman et al., 2024) with a basic channel dimension of 64. By default, a single model with conditional category labels is trained per dataset on normal images for 2,000 epochs using the StableAdamW optimizer with AMSGrad, a learning rate of 10^{-4}, and a batch size of 32.

Parameters and computational cost. Table 4 lists the reference configuration, chosen by sweeping on the test split of MVTec-AD and used unchanged on VisA. A cycle takes $J + 1$ steps, $0 \to \tau_1 \to \cdots \to \tau_J \to 1$, each one network call on K_c states at the first step and on $K_c + K_p$ at each probe time. For the reference ($C = 2$, $J = 3$, $K_p = 32$, $K_c = 4$) that is

$$\underbrace{C(J+1) = 8}_{\text{network calls}}, \qquad \underbrace{C\big[(J+1)K_c + JK_p\big] = 224}_{\text{state evaluations}}, \qquad \underbrace{K_c + K_p = 36}_{\text{largest batch}}, \tag{22}$$

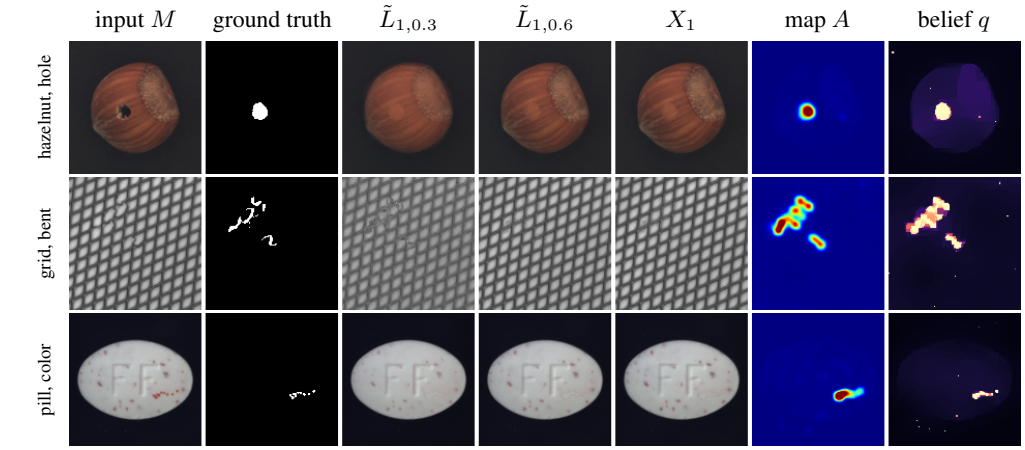


Figure 2: Restorations of three MVTec-AD test images in the last cycle: endpoints $\tilde{L}_{1,t}$ at $t = 0.3$ and 0.6, restoration X_1, anomaly map A and final belief q (bright: rejected). Where the K_c paths differ, each pixel shows the path with the largest residual, as in the readout (21).

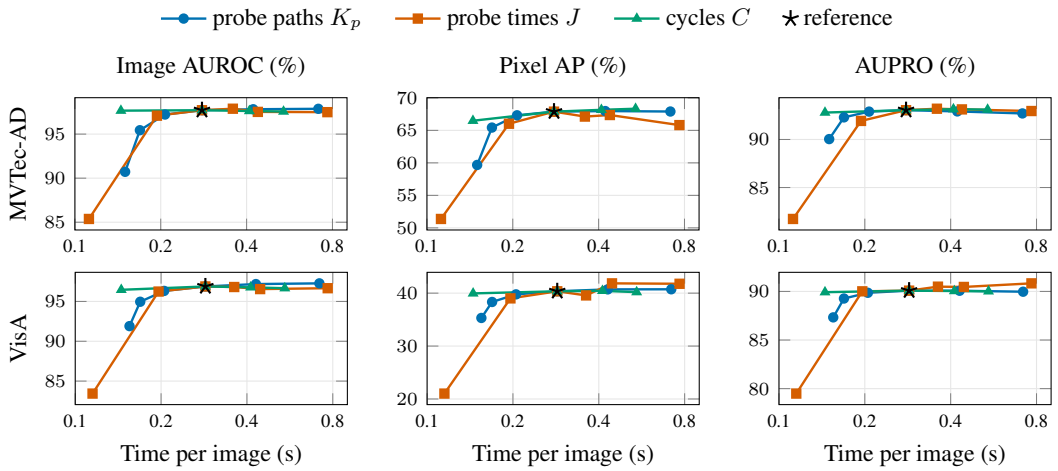


Figure 3: Accuracy against controller time per image (log scale) on MVTec-AD and VisA, mean over seeds 0–2, varying one budget setting at a time: probe paths $K_p \in \{4, 8, 16, 32, 64, 128\}$, probe times $J \in \{1, 2, 3, 4, 5, 9\}$ ending at 0.9, and cycles $C \in \{1, 2, 3, 4\}$. The star is the reference ($K_p = 32$, $J = 3$, $C = 2$). The standard deviation over the seeds is at most 0.44 points (Table 17).

plus one belief solve (7) per measurement, with 100 warm-started FISTA iterations. The controller takes 0.28 s per image on a single NVIDIA A100 GPU on both datasets, and smoothing, scoring and saving the maps add 0.03 s; all times below are controller times.

4.2 Main result

Table 1 compares FlowCTRL with RD4AD (Deng & Li, 2022), CFLOW-AD (Gudovskiy et al., 2022), UniAD (You et al., 2022), RD++ (Tien et al., 2023), SimpleNet (Liu et al., 2023c), DeSTSeg (Zhang et al., 2023), HVQ-Trans (Lu et al., 2023), DiAD (He et al., 2024b), MambaAD (He et al., 2024a), ViTAD (Zhang et al., 2025), InvAD (Zhang et al., 2026), AMDM and OmiAD (Feng et al., 2025), Flow Mismatching (Chen et al., 2026), Dinomaly (Guo et al., 2025) and INP-Former (Luo et al., 2025). On MVTec-AD, FlowCTRL is strongest at pixel-level localization: its pixel AP, F_1-max and AUROC are exceeded only by Dinomaly and INP-Former. On VisA, with the configuration chosen on MVTec-AD unchanged, its image-level metrics are exceeded only by Flow Mismatching, Dinomaly and INP-Former, and each pixel-level metric by four to eight methods. Like

Table 1: Comparison of mainstream anomaly detection methods on MVTec-AD and VisA. FlowCTRL is the mean over 16 seeds, with a std of at most 0.18 points on every metric (Section B.4).

Method	MVTec-AD							VisA						
	Image-level			Pixel-level				Image-level			Pixel-level			
	AUROC	AP	F_1-max	AUROC	AP	F_1-max	PRO	AUROC	AP	F_1-max	AUROC	AP	F_1-max	PRO
RD4AD	94.6	96.5	95.2	96.1	48.6	53.8	91.1	92.4	92.4	89.6	98.1	38.0	42.6	91.8
CFLOW-AD	92.7	97.2	94.0	95.8	46.8	49.6	89.0	87.2	89.5	85.1	97.8	34.2	37.2	87.3
UniAD	96.5	98.8	96.2	96.8	43.4	49.5	90.7	88.8	90.8	85.8	98.3	33.7	39.0	85.5
RD++	95.8	98.0	96.6	97.3	53.0	57.0	92.9	93.1	94.1	90.0	98.4	40.4	44.8	91.4
SimpleNet	95.3	98.4	95.8	96.9	45.9	49.7	86.5	87.2	87.0	81.8	96.8	34.7	37.8	81.4
DeSTSeg	89.2	95.5	91.6	93.1	54.3	50.9	64.8	88.9	89.0	85.2	96.1	39.6	43.4	67.4
HVQ-Trans	98.0	99.5	97.5	97.3	48.2	53.3	91.4	93.2	92.8	87.6	98.7	35.0	39.6	86.3
DiAD	97.2	99.0	96.5	96.8	52.6	55.5	90.7	86.8	88.3	85.1	96.0	26.1	33.0	75.2
MambaAD	98.6	99.6	97.8	97.7	56.3	59.2	93.1	94.3	94.5	89.4	98.5	39.4	44.0	91.0
ViTAD	98.3	99.4	97.3	97.7	55.3	58.7	91.4	90.5	91.7	86.3	98.2	36.6	41.1	85.1
InvAD	98.9	99.6	98.2	98.1	57.1	59.6	94.4	95.6	96.0	92.3	99.0	43.7	46.9	93.0
AMDM	98.4	99.0	97.4	97.5	51.5	56.1	92.6	94.8	95.6	91.1	98.8	39.8	43.5	88.4
OmiAD	98.8	99.7	98.5	97.7	52.6	56.7	93.2	95.3	96.0	91.2	98.9	40.4	44.1	89.2
Mismatching	98.7	99.5	97.7	97.7	67.5	64.8	91.9	98.0	98.4	95.2	97.5	40.2	45.8	90.6
Dinomaly	99.6	99.8	99.0	98.4	69.3	69.2	94.8	98.7	98.9	96.2	98.7	53.2	55.7	94.5
INP-Former	99.7	99.9	99.2	98.5	71.0	69.7	94.9	98.9	99.0	96.6	98.9	51.2	54.7	94.4
FlowCTRL	97.7	99.0	96.6	98.1	67.9	65.2	93.0	96.9	97.2	94.3	98.7	40.3	46.2	90.1

Table 2: Ablation of the components (%), mean over seeds 0–2 and the categories. Recall and FRR: shares of the defect and of the non-defect pixels with final belief $q > \frac{1}{2}$ (Section B.3). s/img: controller time on MVTec-AD (VisA within 0.01 s). Above the middle rule, each row changes one setting; below it, the first row of a block changes only the block's setting (one cycle; a uniform 10-step grid) and the second one more. Detailed numbers with standard deviations in Table 16.

		MVTec-AD					VisA					
		Image	Pixel		Belief		Image	Pixel		Belief		
Part	Setting	AUROC	AP	AUPRO	recall	FRR	AUROC	AP	AUPRO	recall	FRR	s/img
FlowCTRL (reference)		97.7	67.9	93.1	75.0	3.3	96.9	40.3	90.1	86.5	4.2	0.28
Belief	no coupling ($\mu = 0$)	95.0	61.3	91.1	64.9	8.6	92.7	37.9	87.8	79.4	12.5	0.23
Controller	blend at the last step	96.8	62.9	92.0	70.0	2.0	96.9	37.2	91.7	80.8	2.9	0.28
Observer	mean energy over sources	76.6	33.2	74.8	97.8	57.8	72.0	15.5	71.4	99.3	58.8	0.28
	last probe time only	95.1	65.3	89.8	72.3	4.2	96.3	39.4	89.8	88.5	9.3	0.27
Cycles	mean merge	96.5	65.1	92.7	81.4	5.1	95.5	37.9	89.8	92.5	6.2	0.28
One cycle	with control	97.7	66.5	92.8	71.4	2.6	96.5	39.9	89.9	81.6	3.5	0.15
	no control	64.1	20.4	50.3	71.4	2.6	62.6	7.5	55.9	81.5	3.5	0.15
10-step grid	energy on probe paths	97.8	65.9	92.8	72.3	2.4	97.0	39.6	90.9	82.1	3.2	0.38
	energy on controlled path	71.6	27.3	67.6	80.1	23.8	65.0	8.9	61.5	92.1	24.9	0.34

Flow Mismatching and unlike every other baseline, it uses no pretrained network and works in pixel space at 128×128 (Section B.6). Figure 2 shows examples on the restoration.

4.3 Ablation study

Protocol. Every ablation run changes one setting of the reference configuration (Table 4) and is made on both datasets with seed 0; Table 2 and Figure 3 average seeds 0, 1 and 2. Sections B.2, B.3 and B.5 list the tested settings and report all runs and the training checkpoints. The reference settings were chosen on MVTec-AD; on VisA, the highest pixel AP among these runs, 43.7 against 40.3, is obtained with the smoothing $\sigma = 1$ instead of 3. **Components.** Table 2 replaces one component of a part at a time, removes the control at one cycle, and measures the energy on the controlled path instead of the probe paths. Every row lowers pixel AP against the reference on both datasets. **Test-time budget.** Figure 3 varies K_p, J and C, which set the cost with K_c (Equation (22)). Fewer probe paths, probe times or cycles than in the reference lower image AUROC, pixel AP and AUPRO on both datasets; more probe paths or cycles change each by at most 0.5 points, at up to $2.6\times$ the time per image.

AI USE STATEMENT

In this work, we used generative AI tools (large language model assistants, including Claude by Anthropic, Codex by OpenAI) for tasks with required disclosure: to help polish writing including forming better formulation and proofs, to implement the method and its evaluation, to help design the experiments and ablations and propose hypotheses about the role of individual components, and to help interpret the results. We have not used generative AI tools to generate synthetic data sets or to clean or reformat datasets, and translation and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to draft and edit parts of the paper and suggest its structure, to suggest experimental parameters, to write the scripts that build the tables and figures from the saved results, to search for related literature, and to format the references. We have reviewed all AI-assisted work. The results of FlowCTRL in the tables and figures are generated from the saved result files by scripts that also check the comparisons stated in Section 4 and Section B, and every reference was checked, with AI-assisted web search, against its arXiv record and the page of its venue. We manually checked the proofs and tested the code. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

FlowCTRL needs only a flow model trained on normal images and forward passes of it. Section 3 specifies every step of a run, Algorithm 1 (Section A.1) lists the steps in order, and Table 4 gives every setting of the reference configuration in the same notation. Section 4.1 describes the two public datasets, MVTec-AD and VisA, their preprocessing, the architecture and training of the flow model, the metrics, and the computation of a run (22). The results of FlowCTRL are means over 16 seeds of the Gaussian sources, reported with their standard deviations and per category (Section B.4); the settings and components of the method are varied one at a time in Sections B.2 and B.3, and the checkpoints saved during training are evaluated in Section B.5. Table 4 also reports that running in float32 or without compilation changes each metric by less than 0.15 points. The theoretical results are stated with their assumptions and proved in Section A.

REFERENCES

Farzad Beizaee, Sina Hajimiri, Ismail Ben Ayed, Gregory Lodygensky, Christian Desrosiers, and Jose Dolz. REFLECT: Rectified flows for efficient brain anomaly correction transport. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2025.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: differentiating through flows for controlled generation. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 3462–3483, 2024.

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. MVTec AD – a comprehensive real-world dataset for unsupervised anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

Shengzhe Chen, Mehrdad Moradi, Kamran Paynabar, and Hao Yan. Flow mismatching: Unsupervised anomaly detection via velocity discrepancies in flow matching models. *arXiv preprint arXiv:2605.23070*, 2026.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: A patch distribution modeling framework for anomaly detection and localization. In *Pattern Recognition. ICPR International Workshops and Challenges*, pp. 475–489, 2021.

Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9737–9746, 2022.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *International Conference on Learning Representations (ICLR)*, 2025.

Yaoxuan Feng, Wenchao Chen, Yuxin Li, Bo Chen, Yubiao Wang, Zixuan Zhao, Hongwei Liu, and Mingyuan Zhou. Omiad: One-step adaptive masked diffusion model for multi-class anomaly detection via adversarial distillation. In *Forty-second International Conference on Machine Learning*, pp. 16604–16633, 2025.

Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 98–107, 2022.

Jia Guo, Shuai Lu, Weihang Zhang, Fang Chen, Huiqi Li, and Hongen Liao. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 20405–20415, 2025.

Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. MambaAD: Exploring state space models for multi-class unsupervised anomaly detection. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024a.

Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 8472–8480, 2024b.

Liangwei Li, Lin Liu, Hanzhe Liang, Juanxiu Liu, Jing Zhang, Ruqian Hao, Xiaohui Du, Yong Liu, and Pan Li. Time-reversed flow matching with worst transport in high-dimensional latent space for image anomaly detection. *arXiv preprint arXiv:2508.05461*, 2025a.

Zhong Li, Qi Huang, Yuxuan Zhu, Lincen Yang, Mohammad Mohammadi Amiri, Niki van Stein, and Matthijs van Leeuwen. Scalable, explainable and provably robust anomaly detection with one-step flow matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025b.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code, 2024. URL `https://arxiv.org/abs/2412.06264`.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023a.

Xingchao Liu, Lemeng Wu, Shujian Zhang, Chengyue Gong, Wei Ping, and Qiang Liu. Flowgrad: Controlling the output of generative odes with gradients. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24335–24344, 2023b.

Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20402–20411, 2023c.

Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pp. 8487–8500, 2023.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using denoising diffusion probabilistic models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11461–11471, 2022.

Wei Luo, Yunkang Cao, Haiming Yao, Xiaotian Zhang, Jianan Lou, Yuqi Cheng, Weiming Shen, and Wenyong Yu. Exploring intrinsic normal prototypes within a single image for universal anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9974–9983, 2025.

Ségolène Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. PnP-Flow: Plug-and-play image restoration with flow matching. In *International Conference on Learning Representations (ICLR)*, 2025.

Ashwini Pokle, Matthew J. Muckley, Ricky T. Q. Chen, and Brian Karrer. Training-free linear image inverses via flows. *Transactions on Machine Learning Research*, 2024.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Litu Rout, Yujia Chen, Nataniel Ruiz, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Semantic image inversion and editing using rectified stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2025.

Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan T. M. Duong, Chanh D. Tr. Nguyen, and Steven Q. H. Truong. Revisiting reverse distillation for anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24511–24520, 2023.

Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. DRAEM – a discriminatively trained reconstruction embedding for surface anomaly detection. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Jiangning Zhang, Xuhai Chen, Yabiao Wang, Chengjie Wang, Yong Liu, Xiangtai Li, Ming-Hsuan Yang, and Dacheng Tao. Exploring plain ViT features for multi-class unsupervised visual anomaly detection. *Computer Vision and Image Understanding*, 253:104308, 2025.

Jiangning Zhang, Chengjie Wang, Xiangtai Li, Guanzhong Tian, Zhucun Xue, Yong Liu, Guansong Pang, and Dacheng Tao. Learning feature inversion for multi-class anomaly detection under general-purpose COCO-AD benchmark. *International Journal of Computer Vision*, 134(5):235, 2026. doi: 10.1007/s11263-026-02809-z.

Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3914–3923. IEEE, 2023.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European Conference on Computer Vision*, pp. 392–408, 2022.

APPENDIX A DETAILED THEORIES

This section completes the steps that Section 3 states without proof, in the order of the method: the joint problem of Section 3.1 (Section A.2), the observer of Section 3.2 (Section A.3), the controller of Section 3.3 (Section A.4), and the cycles and the readout of Section 3.4 (Section A.5). Section A.1 first writes one run as an algorithm. Throughout, $n = HW$ is the number of pixels; maps without a channel axis, such as the energy d, the evidence S, the belief q and the weight w, live in $\mathbb{R}^{H\times W}$ and act on images by broadcasting across the channels; s is the sensor scale of (1), and $s_t = 1 - t$ is the scale that a state at time t still carries.

Algorithm 1 One run of FlowCTRL. Colours: observer, controller. The operators are (23)– (25).

Require: $M \in \Omega$; velocity v_θ; cycles C; probe times $\tau_1 < \cdots < \tau_J$; sources K_p, K_c; belief parameters (λ, T, μ); filter rate κ

$Y \leftarrow M, \quad \mathcal{S} \leftarrow \varnothing$ ▷ probe target, and the finished cycles' evidence

for $c = 1, \dots, C$ **do**

  $\boldsymbol{U} \sim \mathcal{N}(0, I)^{\max(K_p, K_c)}, \quad \boldsymbol{X} \leftarrow \boldsymbol{U}_{1:K_c}, \quad S \leftarrow \perp, \quad q \leftarrow \perp$ ▷ path k shares $\boldsymbol{U}_k$

  for $j = 0, \dots, J$ **do**

    $t \leftarrow t_j, \quad t' \leftarrow t_{j+1}, \quad s \leftarrow 1 - t$ ▷ grid $t_0 = 0,\ t_j = \tau_j,\ t_{J+1} = 1$

    $\boldsymbol{P} \leftarrow (1 - \tau_j)\, \boldsymbol{U}_{1:K_p} + \tau_j\, Y$ if $j \geq 1$, else $\varnothing$ ▷ Equation (9)

    $(\boldsymbol{V}_X, \boldsymbol{V}_P) \leftarrow v_\theta\big([\boldsymbol{X}; \boldsymbol{P}],\, [t\mathbf{1}_{K_c};\, \tau_j \mathbf{1}_{K_p}]\big)$ ▷ one batched call, per-sample times

    $\hat{L} \leftarrow \boldsymbol{X} + s\, \boldsymbol{V}_X, \quad \hat{U} \leftarrow \boldsymbol{X} - t\, \boldsymbol{V}_X$ ▷ Equation (18)

    if $j \geq 1$ **then**

      $d \leftarrow \mathrm{smin}_T\, \mathcal{E}_{1-\tau_j}\big(M - (\boldsymbol{P} + (1 - \tau_j)\boldsymbol{V}_P)\big)$ ▷ always against M, Equation (10)

      $S \leftarrow (1 - \gamma_j)\, S + \gamma_j\, d, \quad \gamma_j = 1 - e^{-\kappa(\tau_j - \tau_{j-1})}$ $(S \leftarrow d$ if $S = \perp)$ ▷ Equation (11)

      $\bar{S} \leftarrow \mathrm{smin}_T\, \big(\mathcal{S} \cup \{S\}\big), \quad q \leftarrow \mathcal{G}_\mu(\bar{S})$ ▷ Section A.5.1, Equation (7)

      $\tilde{L} \leftarrow \begin{cases} \hat{L}, & q = \perp \quad \text{(no belief yet: Euler)} \\ (1 - q) \odot M + q \odot \hat{L}, & j = J \quad (20) \\ \big(\hat{L} + w \odot M\big)/(1 + w), \quad w = (1 - q)/s^2, & \text{otherwise} \quad (17) \end{cases}$

    $\boldsymbol{X} \leftarrow (1 - t')\, \hat{U} + t'\, \tilde{L}$ ▷ Equation (19)

  $\boldsymbol{X}_1 \leftarrow \boldsymbol{X}, \quad Y \leftarrow \frac{1}{K_c} \sum_k \boldsymbol{X}_{1,k}, \quad \mathcal{S} \leftarrow \mathcal{S} \cup \{S\}$ ▷ Section 3.4

return $A \leftarrow \sqrt{2 \max_k \mathcal{E}_1(M - \boldsymbol{X}_{1,k})}$ ▷ Equation (21)

A.1 One run in vector form

Algorithm 1 is written on whole arrays. Let $\Omega = \mathbb{R}^{C \times H \times W}$ be the space of one image, so that $M, \hat{L}, \hat{U}, \tilde{L}, X_1 \in \Omega$, and let a **bold** symbol be a stack of K of them, an element of Ω^K; K_c controlled states are $\boldsymbol{X}$ and K_p probe states are $\boldsymbol{P}$, and $\odot$ and $/$ broadcast maps across the channel and stack axes. Three reductions are named once and then used without indices:

$$\mathcal{E}_s(R) = \tfrac{1}{2s^2}\, \|R\|_{\mathrm{ch}}^2 \qquad \text{per-pixel energy (3);} \tag{23}$$

$$\mathrm{smin}_T(\boldsymbol{a}) = -T \log \tfrac{1}{K} \textstyle\sum_k e^{-\boldsymbol{a}_k / T} \qquad \text{soft minimum over a stack (10);} \tag{24}$$

$$\mathcal{G}_\mu(S) = \arg\min_q \textstyle\sum_i \psi(S_i, q_i) + \mu \sum_{i \sim j} |q_i - q_j| \qquad \text{coupled belief (7).} \tag{25}$$

The first two map an image, or a stack of them, to one $H \times W$ map; the third maps an $H \times W$ map to a belief in $[0, 1]^{HW}$. $\mathcal{G}_\mu$ is the only one that is not elementwise, and Section A.3.2 is how it is solved. Everything else in the run is broadcasting arithmetic. Figure 4 draws one step of the controller and one probe.

A.2 The joint problem

This subsection concerns (3)–(6) for a fixed normal image L, so that $d = d(L)$ is a fixed map. Write the objective of the belief problem (7) and its value as

$$F_T(q; d) = \sum_i \psi(d_i, q_i) + \mu \sum_{i \sim j} |q_i - q_j|, \qquad \Phi_\mu(d) = \min_{q \in [0,1]^n} F_T(q; d), \tag{26}$$

so that $\Phi_0(d) = \sum_i \Phi(d_i)$ by (8), and write $F_0(q; d)$ for F_T with the entropy removed, $\sum_i \big[(1 - q_i)d_i + q_i \lambda\big] + \mu \sum_{i \sim j} |q_i - q_j|$, which on binary $q = z$ is the objective of (3) at the fixed L, up to the effort, which is then a constant. The belief is $q^\star(d) = \arg\min_q F_T(q; d)$.

A.2.1 Relaxing the support, and the decision at one half

Proposition 1 (relaxation). *For every map d:*

(i) $\min_{q \in [0,1]^n} F_0(q; d) = \min_{z \in \{0,1\}^n} F_0(z; d)$;

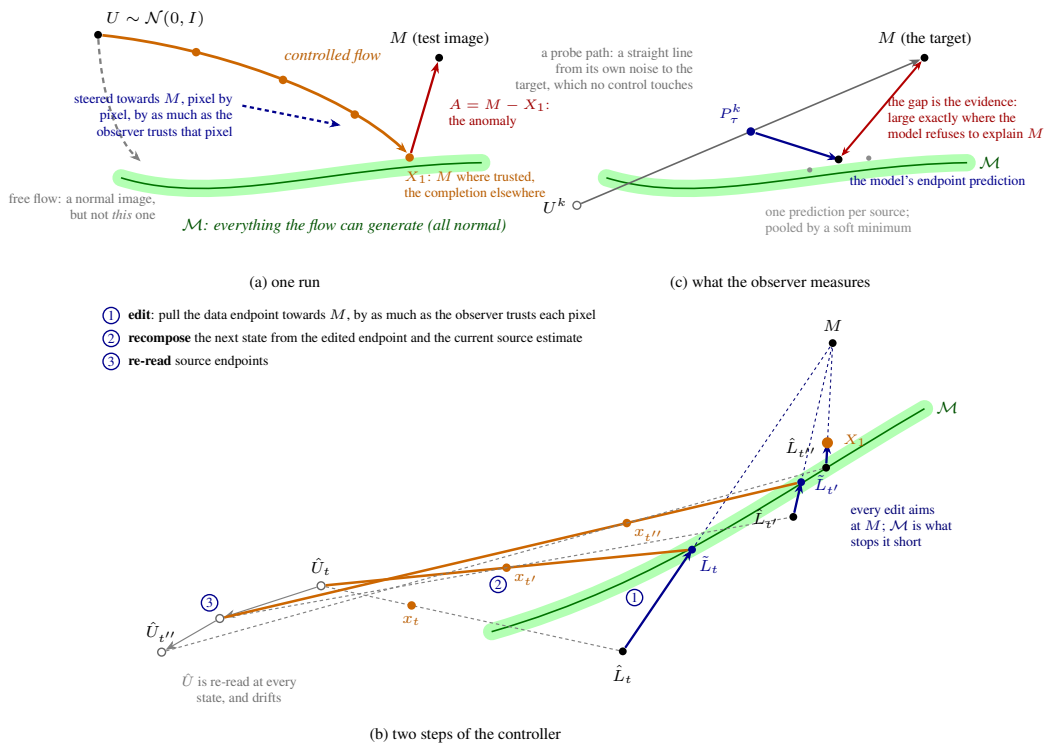


Figure 4: **Detection as a control problem** (schematic). **(a)** A flow model trained on normal images produces normal images, so running it from a Gaussian draw gives a normal image unrelated to the test image M (dashed). We *steer* the same trajectory instead: it ends at a restoration X_1 that keeps M where the belief trusts a pixel and takes the model's own completion where it does not, and $A = M - X_1$ is the anomaly map. **(b)** How a step works, drawn twice. A state is a point on the chord between the model's two endpoint estimates, the source $\hat{U}$ and the data endpoint $\hat{L}$ (18). The controller edits only the data endpoint, pulling it towards M by as much as the belief trusts each pixel (17), and recomposes the next state on the new chord; at that state *both* endpoints are read again. $\hat{L}$ is an estimate: early on it is a blur of everything the path could still become, and it sharpens as t grows. The source estimate is re-read as well and drifts away from the draw the path started at. The manifold $\mathcal{M}$ is drawn only as a picture of the normal images. **(c)** Where the belief comes from. A *probe* path is a straight line from its own noise to the target (9), which no control ever touches. The model predicts an endpoint at a state on it, and the gap between that prediction and M is the evidence (10); it is large where the model does not explain the observation. One prediction per source is pooled into one map, from which the per-pixel belief that drives (b) is computed.

(ii) $0 \le \min_{z\in\{0,1\}^n} F_0(z;d) - \Phi_\mu(d) \le nT\log 2$, *with equality on the right when* $d_i = \lambda$ *for every* i*;*

(iii) the belief $q^\star(d)$ *is unique and lies in* $(0,1)^n$*, and for every* $\theta \in (0,1)$ *the binary map* $z^\theta = \mathbf{1}\{q^\star(d) > \theta\}$ *minimises* $F_0(\cdot\,;d)$ *over* $\{0,1\}^n$ *with* λ *replaced by* $\lambda + T\log\frac{\theta}{1-\theta}$*. In particular* $\mathbf{1}\{q^\star(d) > \frac{1}{2}\}$ *minimises the binary problem at the original* λ*, for every* $T > 0$*.*

Proof. (i) Every binary z lies in the cube, so the left side is at most the right. Conversely, fix $q \in [0,1]^n$ and put $z_i^\theta = \mathbf{1}\{q_i > \theta\}$ for $\theta \in (0,1)$. Then $q_i = \int_0^1 z_i^\theta \,\mathrm{d}\theta$ and $|q_i - q_j| = \int_0^1 |z_i^\theta - z_j^\theta|\,\mathrm{d}\theta$, because for every θ between q_i and q_j exactly one of the two indicators is one. Since $F_0(\cdot\,;d)$ is affine in q apart from the coupling, $F_0(q;d) = \int_0^1 F_0(z^\theta;d)\,\mathrm{d}\theta \ge \min_z F_0(z;d)$.

(ii) With $h(u) = u\log u + (1-u)\log(1-u)$ we have $-\log 2 \le h \le 0$ on $[0,1]$, so $F_0 - nT\log 2 \le F_T \le F_0$ at every q. Taking minima and using (i) gives both bounds. At $d_i = \lambda$ for all i every binary

z costs $n\lambda$ plus a nonnegative coupling, and the constant map $q = \frac{1}{2}\mathbf{1}$ has $F_T = n\lambda - nT\log 2$, so the right bound is attained.

(iii) $h''(u) = 1/(u(1-u)) \geq 4$, so $F_T(\cdot\,; d)$ is $4T$-strongly convex and has exactly one minimiser. The derivative $T\log\frac{q_i}{1-q_i}$ of the entropy diverges at 0 and 1, while every subgradient of the coupling is bounded, so the minimiser is interior. With D the difference operator of the grid (Section A.3.2), optimality there reads, for some p with $p_e \in \operatorname{sign}\big((Dq^\star)_e\big)$ and $|p_e| \leq 1$,

$$0 = \lambda\mathbf{1} - d + T\log\frac{q^\star}{1-q^\star} + \mu D^\top p.$$

Fix θ and put $\lambda_\theta = \lambda + T\log\frac{\theta}{1-\theta}$. On an edge that the level set z^θ cuts, Dz^θ and $Dq^\star$ have the same sign, and on an edge it does not cut every $p_e \in [-1,1]$ is a subgradient, so $\mu D^\top p \in \mu\,\partial\|Dz^\theta\|_1$. The vector $g = \lambda_\theta\mathbf{1} - d + \mu D^\top p = T\big(\log\frac{\theta}{1-\theta} - \log\frac{q^\star}{1-q^\star}\big)$ is then a subgradient at z^θ of the convex function F_0 at price λ_θ, with $g_i < 0$ where $z_i^\theta = 1$ and $g_i \geq 0$ where $z_i^\theta = 0$. That is the optimality condition of minimising F_0 over the cube at z^θ, and by (i) z^θ minimises the binary problem as well. At $\theta = \frac{1}{2}$, $\lambda_\theta = \lambda$. □

Part (i) is the remark of Section 3.1 that relaxing z to the cube changes nothing by itself. Part (ii) bounds what the entropy changes in value, at most $T\log 2$ per pixel; (30) is its uncoupled, per-pixel form. Part (iii) says that the entropy changes nothing in the decision taken at one half: the pixels whose belief exceeds $\frac{1}{2}$ form an exact minimiser of the binary problem at every temperature, and every other threshold θ solves the same problem at a shifted price. The belief columns of the tables in Section B.3 use this decision.

A.2.2 The two factors of the joint problem

Section 3.1 reads the costs of (3) as probabilities. Take the distribution p_θ of the images that the flow produces from Gaussian draws without control in the role of the prior on L that the effort expresses (the effort is zero exactly on these images), and turn the other costs into probabilities at the temperature T of the belief:

$$p(L, z \mid M) \;\propto\; p_\theta(L)\,\exp\Big\{-\frac{1}{T}\Big[\sum_i \big((1-z_i)\,d_i(L) + z_i\lambda\big) + \mu\sum_{i\sim j}|z_i - z_j|\Big]\Big\}. \tag{27}$$

The product rule $p(z, L \mid M) = p(z \mid M)\,p(L \mid M, z)$ then has the two factors

$$p(z \mid M) \;\propto\; e^{-[\lambda\sum_i z_i + \mu\sum_{i\sim j}|z_i - z_j|]/T}\;\mathbb{E}_{L\sim p_\theta}\,e^{-\sum_i (1-z_i)\,d_i(L)/T}, \tag{28}$$

$$p(L \mid M, z) \;\propto\; p_\theta(L)\,e^{-\sum_i (1-z_i)\,d_i(L)/T}. \tag{29}$$

The first factor needs only the images of the flow without control and the observation: it scores a set of anomalous pixels by how well those images explain the other pixels, on average. The second factor is the flow's images weighted by how well they agree with M on the pixels that z trusts, which is the target that the control problem (12) steers towards. For one pixel considered alone ($\mu = 0$), (28) gives the probability $\sigma\big((\bar{d}_i - \lambda)/T\big)$ that the pixel is anomalous, with $\bar{d}_i = -T\log\mathbb{E}_{L\sim p_\theta}e^{-d_i(L)/T}$; the observer replaces the images of the flow by its predictions on the probe paths and estimates this average over their sources (Section A.3.4).

A.2.3 Why the belief is not read off the path being steered

The belief could be read off the path being steered, for free, since $\hat{L}_t(x_t)$ is already computed at every step. Section 3.1 says that solving the two blocks of (6) in turn this way fails. The following is the mechanism, stated at a frozen time so that the feedback is isolated from the motion of the path.

Lemma 1 (self-erasure)**.** *Fix t, write $s = s_t$, and suppose the belief is read off the controlled endpoint itself: $q^{(k)} = \sigma\big((d^{(k)} - \lambda)/T\big)$ with $d^{(k)} = \|M - L^{(k)}\|^2/(2s^2)$. Apply the blend* (17) *repeatedly to one pixel,*

$$L^{(k+1)} = \frac{L^{(k)} + w^{(k)}M}{1 + w^{(k)}}, \qquad w^{(k)} = \frac{1 - q^{(k)}}{s^2} \qquad (\mu = 0).$$

Then $w^{(k)} \geq w^{(0)} > 0$ *for every* k*,*

$$\big\|M - L^{(k)}\big\| \;\leq\; \big(1 + w^{(0)}\big)^{-k} \big\|M - L^{(0)}\big\| \;\longrightarrow\; 0,$$

and $q^{(k)} \to \sigma(-\lambda/T)$*.*

Proof. By (48) the residual obeys $M - L^{(k+1)} = (M - L^{(k)})/(1 + w^{(k)})$, so it is non-increasing in norm whenever $w^{(k)} \geq 0$. The energy d is increasing in that norm and the belief is increasing in d, so $q^{(k)}$ is non-increasing, hence $w^{(k)} = (1 - q^{(k)})/s^2$ is non-decreasing and $w^{(k)} \geq w^{(0)}$; and $w^{(0)} > 0$ because $q^\star < 1$ for every finite residual. Iterating the residual recursion gives the geometric bound, and $d^{(k)} \to 0$ gives $q^{(k)} \to \sigma((0 - \lambda)/T)$. □

The loop is therefore monotone in the wrong direction, and it is fast. At the defaults of Section A.3.7 a pixel carrying no evidence has $q = \sigma(-4) = 0.018$, so at $t = 0.6$ ($s = 0.4$) its weight is $w = 6.14$: one step divides its residual by 7.1 and two steps by 51. A pixel that starts clearly rejected barely moves, because $w \approx 0$ there. The pixels at risk are exactly the anomalous ones whose residual has not *yet* crossed the threshold: the controller absorbs precisely the anomalies it has not yet noticed, and in doing so destroys the evidence it would have needed to notice them. The probe paths of (9) avoid the loop by construction: no control ever touches them, so within a cycle the belief does not depend on the controlled states at all (Proposition 7). The cost of that separation is the K_p extra states per probe time, which ride in the same network call as the controlled states.

A.3 The observer

A.3.1 The belief in closed form

This is (8), with the properties the rest of the method uses. Everything is per pixel until the last paragraph, so fix one pixel and drop the index.

The minimiser. Differentiating the free energy (5) in q,

$$\partial_q \psi(d, q) \;=\; -d + \lambda + T\big[\log q + 1 - \log(1 - q) - 1\big] \;=\; -d + \lambda + T \log \frac{q}{1 - q},$$

which vanishes exactly at $q^\star = \sigma\big((d - \lambda)/T\big)$, and $\partial_q^2 \psi = T\big[\frac{1}{q} + \frac{1}{1-q}\big] = \frac{T}{q(1-q)} > 0$ on $(0, 1)$, so ψ is strictly convex and the stationary point is the unique minimum.

The value is a soft minimum. Write $c = (d - \lambda)/T$, so $q^\star = \sigma(c)$, $\log q^\star = c - \operatorname{sp}(c)$ and $\log(1 - q^\star) = -\operatorname{sp}(c)$ with $\operatorname{sp}(c) = \log(1 + e^c)$. Substituting into (5),

$$\psi(d, q^\star) = (1-q^\star)d + q^\star\lambda + T\big[q^\star c - \operatorname{sp}(c)\big] = (1-q^\star)d + q^\star\lambda + q^\star(d-\lambda) - T\operatorname{sp}(c) = d - T\operatorname{sp}(c),$$

and pulling $d = -T \log e^{-d/T}$ inside the logarithm gives $\Phi(d) = -T \log\big(e^{-d/T} + e^{-\lambda/T}\big)$, which is (8). Since $\max(a, b) \leq \log(e^a + e^b) \leq \max(a, b) + \log 2$ with $a = -d/T$ and $b = -\lambda/T$,

$$0 \;\leq\; \min\{d, \lambda\} - \Phi(d) \;\leq\; T \log 2, \tag{30}$$

so Φ is the truncated quadratic of (4) to within $T \log 2$, and recovers it as $T \to 0$. The entropy relaxation changes the decision and not the cost; Proposition 1 is the coupled version.

A pixel with no evidence is not fully trusted. At $d = 0$ the belief is $q^\star = \sigma(-\lambda/T)$, which at the defaults of Section A.3.7 is $\sigma(-4) \approx 0.018$. The controller therefore never pulls a pixel all the way to M before the last step, and the blend weight w of (17) is bounded by $0.982/s_t^2$.

The influence of a pixel is $1 - q^\star$. ψ is affine in d with slope $1 - q$, so by the envelope theorem $\Phi'(d) = 1 - q^\star(d)$. The same holds with the coupling: $F_T(q; d)$ of (26) is affine in d with slope $\mathbf{1} - q$ for every q, and its minimiser is unique (Proposition 1), so by Danskin's theorem the value Φ_μ is concave and differentiable, with

$$\nabla_d \Phi_\mu(d) \;=\; \mathbf{1} - q^\star(d), \qquad -\nabla_L\, \Phi_\mu\big(d(L)\big) \;=\; \big(\mathbf{1} - q^\star(d(L))\big) \odot \frac{M - L}{s^2}, \tag{31}$$

the second by the chain rule with $d_i(L) = \|M_i - L_i\|^2/(2s^2)$. At $\mu = 0$ the first reads $\Phi'(d_i) = 1 - q_i^\star(d_i)$. A rejected pixel has $q^\star \to 1$ and therefore contributes no pull on L: the belief is exactly the factor by which a pixel's residual is allowed to act on anything downstream, which is why the same $1 - q$ appears in the controller's weight (17) and in the stationarity condition of Section A.4.2 without having to be put there.

A.3.2 Computing the coupled belief

Equation (7) is what Section 3.2 defers. It is strictly convex on a compact set, so the minimiser exists and is unique, but it is not separable and has no closed form. We solve it in the dual.

In logit units. Drop the constant $\sum_i d_i$ from $\sum_i \psi(d_i, q_i)$ and divide by T. With the logits $\ell_i = (d_i - \lambda)/T$, the coupling weight $\nu = \mu/T$ and the forward-difference operator D of the 4-neighbour grid, so that $\sum_{i\sim j} |q_i - q_j| = \|Dq\|_1$, the problem (7) reads

$$\min_{q\in[0,1]^{HW}} \underbrace{\sum_i \big[-q_i\ell_i + q_i \log q_i + (1-q_i)\log(1-q_i)\big]}_{=:\,F(q)} + \nu\,\|Dq\|_1. \tag{32}$$

At $\nu = 0$ each coordinate gives $q_i = \sigma(\ell_i)$, which is (8).

The dual. Write $\nu\|Dq\|_1 = \max_{\|p\|_\infty \le \nu} \langle p, Dq\rangle$ over one variable p_e per edge. F is strictly convex and the box is compact, so the minimum and the maximum may be exchanged. For fixed p the inner problem is separable, with $c_i = \ell_i - (D^\top p)_i$:

$$\min_{q_i} \; -q_i c_i + q_i \log q_i + (1-q_i)\log(1-q_i) \quad \Longrightarrow \quad q_i(p) = \sigma(c_i),$$

whose value, by the same substitution as in Section A.3.1, is $-\operatorname{sp}(c_i)$. The dual is therefore the smooth box-constrained problem

$$\min_{\|p\|_\infty \le \nu} G(p) := \sum_i \operatorname{sp}\big(\ell_i - (D^\top p)_i\big), \qquad \nabla G(p) = -D\,q(p), \qquad q(p) = \sigma\big(\ell - D^\top p\big). \tag{33}$$

Its Hessian is $D\,\mathrm{diag}(\sigma')D^\top$ with $\sigma' \le \frac{1}{4}$ and $\|D\|^2 \le 8$ on the 4-neighbour grid, so ∇G is 2-Lipschitz and projected gradient with step $\frac{1}{2}$ converges; we use its FISTA acceleration, 100 iterations, warm-started from the previous probe time and the previous cycle, since S changes little between measurements.

The returned pair is always feasible, and its gap is reported. The belief returned is the primal recovery $q(p)$ of the last iterate, so it lies in $[0,1]^{HW}$ whatever the iteration budget. For that pair,

$$\text{primal} - \text{dual} \;=\; \nu\|Dq\|_1 - \langle p, Dq\rangle \;\ge\; 0 \tag{34}$$

by Hölder's inequality, and it is zero exactly at the optimum. The implementation logs it per pixel in energy units.

What the coupling does, read at $T \to 0$. In the zero-temperature limit (32) becomes the binary problem of (3), whose relaxation is exact (Proposition 1), and the coupling is a penalty on the boundary length of the rejected set. Let $R = \{i : z_i = 1\}$ and let $|\partial R|$ be the number of neighbour pairs with exactly one endpoint in R. Against $z \equiv 0$ the objective changes by $\sum_{i\in R}(\lambda - S_i) + \mu|\partial R|$, so R is rejected exactly when

$$\sum_{i\in R}(S_i - \lambda) \;>\; \mu\,|\partial R|. \tag{35}$$

An isolated pixel has $|\partial R| = 4$ and so needs $S_i > \lambda + 4\mu$; a one-pixel-wide line of length ℓ has $|\partial R| = 2\ell + 2$ and needs a mean excess above $\mu(2 + 2/\ell)$; an $a \times a$ square has $|\partial R| = 4a$ over a^2 pixels and needs a mean excess above $4\mu/a$. At the defaults ($\lambda = 0.08$, $\mu = 0.08$) an isolated pixel needs $S > 0.40$, five times the threshold, while a large region needs little more than λ: the coupling trades single-pixel sensitivity for contiguity.

A.3.3 THE BELIEF IS A STABLE FUNCTION OF THE EVIDENCE

Proposition 2 (stability)**.** *For all maps d and d',*

$$\big\|q^\star(d) - q^\star(d')\big\|_2 \;\le\; \frac{1}{4T}\,\|d - d'\|_2,$$

and $\frac{1}{4T}$ is the smallest constant that holds for all maps, at every $\mu \ge 0$.

Proof. $F_T(q;d) = \sum_i d_i + \varphi(q) - d^\top q$ with $\varphi(q) = \lambda \mathbf{1}^\top q + T\sum_i h(q_i) + \mu\|Dq\|_1$ restricted to the cube, so $q^\star(d)$ minimises $\varphi(q) - d^\top q$, that is, $d \in \partial\varphi\big(q^\star(d)\big)$. As in Proposition 1, φ is $4T$-strongly convex, so its subdifferential is $4T$-strongly monotone: $\langle d - d', q^\star(d) - q^\star(d')\rangle \ge 4T\|q^\star(d) - q^\star(d')\|_2^2$, and the Cauchy–Schwarz inequality gives the bound. Along constant maps $d = \bar{d}\mathbf{1}$ the minimiser is the constant belief $\sigma\big((\bar{d} - \lambda)/T\big)\mathbf{1}$, which satisfies the optimality condition of Proposition 1 with $p = 0$, and its derivative in $\bar{d}$ at $\bar{d} = \lambda$ is $\frac{1}{4T}$ per pixel. □

The belief moves by at most $\frac{1}{4T}$ times the change of the evidence that produces it, whatever the coupling. The temperature therefore sets at once how sharp the decision is and how strongly it reacts to the evidence: at $T = 0.02$ the constant is 12.5 per unit of energy, and it is reached by pixels whose evidence sits at the threshold λ.

A.3.4 THE SOFT MINIMUM OVER SOURCES

The soft minimum averages likelihoods over sources. The K_p energies of (10) differ only in the source U^k of their probe. Consider one pixel alone ($\mu = 0$). If the energy of the hypothesis that the pixel is normal is $d_i(U)$ for a source U, the likelihood of that hypothesis averaged over sources is $\mathbb{E}_U e^{-d_i(U)/T}$, the likelihood of the anomalous hypothesis is $e^{-\lambda/T}$, and the probability that the pixel is anomalous is

$$\sigma\Big(-\frac{\lambda}{T} - \log \mathbb{E}_U e^{-d_i(U)/T}\Big) \;=\; \sigma\Big(\frac{\bar{d}_i - \lambda}{T}\Big), \qquad \bar{d}_i = -T\log \mathbb{E}_U e^{-d_i(U)/T}.$$

The pooled energy $d_{\tau,i}$ of (10) is the Monte Carlo estimate of $\bar{d}_i$ from the K_p probes, and inserting it into (8) gives the formula of Section 3.2. Averaging the likelihood rather than the energy is what makes the pooling a soft minimum, and the following says how it behaves.

Proposition 3 (pooling bounds)**.** *Let $m = \min_k d^k$ and $\bar{d} = \frac{1}{K}\sum_k d^k$, and let $d_{\text{eff}} = -T\log\frac{1}{K}\sum_k e^{-d^k/T}$. Then*

$$m \;\le\; d_{\text{eff}} \;\le\; m + T\log K, \qquad d_{\text{eff}} \;\le\; \bar{d},$$

with equality in the last exactly when all d^k agree. As $T \to 0$, $d_{\text{eff}} \to m$; as $T \to \infty$, $d_{\text{eff}} \to \bar{d}$.

Proof. $\frac{1}{K}\sum_k e^{-d^k/T} \le e^{-m/T}$ gives the lower bound and $\frac{1}{K}\sum_k e^{-d^k/T} \ge \frac{1}{K}e^{-m/T}$ the upper one. For the third, Jensen's inequality applied to the concave logarithm gives $\log\frac{1}{K}\sum_k e^{-d^k/T} \ge \frac{1}{K}\sum_k(-d^k/T) = -\bar{d}/T$, and multiplying by $-T$ reverses it, with equality iff the $e^{-d^k/T}$ are all equal. The two limits follow from the first display and from $\log\frac{1}{K}\sum_k e^{-d^k/T} = -\bar{d}/T + O(T^{-2})$. □

What this buys is a one-sided reading of disagreement between sources. A pixel counts as *explained* as soon as some plausible source reproduces it, whereas the mean $\bar{d}$ asks that the average source reproduce it. Normal content that the model can complete in several ways – a texture, an edge whose phase is not determined by its surroundings – is reproduced by a minority of the draws, so the mean rejects it and the soft minimum does not.

The estimate also improves with K_p in a way the mean does not need to. Let $u_k = e^{-d^k/T}$ be i.i.d. and positive and let I be a uniformly random subset of $\{1,\dots,K+1\}$ of size K. Every index lies in I with probability $K/(K+1)$, so $\frac{1}{K+1}\sum_{k\le K+1} u_k = \mathbb{E}_I\big[\frac{1}{K}\sum_{k\in I} u_k\big]$, and Jensen gives $\log\frac{1}{K+1}\sum_k u_k \ge \mathbb{E}_I \log\frac{1}{K}\sum_{k\in I} u_k$. Taking expectations and using exchangeability,

$\mathbb{E}\big[d_{\text{eff}}^{(K+1)}\big] \le \mathbb{E}\big[d_{\text{eff}}^{(K)}\big]$: the pooled energy is non-increasing in the number of sources, in expectation, and is bounded below by $-T \log \mathbb{E}u_1$.

The same operator merges the cycles (Section A.5.1); there the pooled maps come from different cycles and are not identically distributed, so only the bounds of Proposition 3 carry over. The readout (21) pools the K_c restorations by their maximum instead (Section A.5.3).

A.3.5 What a probe measures

Proposition 4 (probe identity). *For every velocity field v_θ, time $\tau \in [0,1)$, source U, target Y and observation M, the probe state $P_\tau = (1-\tau)U + \tau Y$ of* (9) *satisfies*

$$M - \hat{L}_\tau(P_\tau) \;=\; s_\tau\big[(Y-U) - v_\theta(P_\tau,\tau)\big] + (M-Y), \tag{36}$$

so the energy of probe k in (10) *is* $d^k_{\tau,i} = \frac{1}{2}\big\|\big[(Y-U^k) - v_\theta(P^k_\tau,\tau)\big]_i + (M_i - Y_i)/s_\tau\big\|^2$.

Proof. $\hat{L}_\tau(P_\tau) = P_\tau + s_\tau v_\theta(P_\tau,\tau)$ and $P_\tau - Y = s_\tau(U-Y)$, so $M - \hat{L}_\tau(P_\tau) = (M-Y) - (P_\tau - Y) - s_\tau v_\theta = (M-Y) + s_\tau(Y-U) - s_\tau v_\theta$. Dividing by s_τ and squaring gives the energy. □

In the first cycle $Y = M$, the offset vanishes, and $d^k_{\tau,i} = \frac{1}{2}\big\|(M-U^k) - v_\theta(P^k_\tau,\tau)\big\|^2_i$: half the squared difference, at pixel i, between the velocity $M - U^k$ of the probe's own path and the velocity the model predicts on it. This is the per-pixel flow-matching loss that M would incur as a training sample with source U^k at time τ, the velocity mismatch; it is small where M looks like the training data, and the scale s_τ of (10) is exactly what turns the endpoint residual into this velocity. From the second cycle on, Y_c averages restorations that end with (20), and averaging $M - X^k_1 = q_{\tau_J} \odot \big(M - \hat{L}^k_{1,\tau_J}\big)$ over the K_c paths of cycle $c-1$ gives

$$M - Y_c \;=\; q^{(c-1)}_{\tau_J} \odot \big(M - \bar{L}^{(c-1)}\big), \tag{37}$$

where $\bar{L}^{(c-1)}$ is the mean of the paths' completions $\hat{L}^k_{1,\tau_J}$ and $q^{(c-1)}_{\tau_J}$ is the final belief of cycle $c-1$. The offset of (36) therefore vanishes where the previous cycle trusted the pixel. On a pixel it rejected, the offset is the gap between the previous completion and the observation, enlarged by $1/s_\tau$ in the energy: a rejected pixel whose completion agrees with M loses the offset and its evidence falls, while one whose completion departs from M keeps it.

A.3.6 The filter is grid-invariant

If d is constant on $[\tau_{j-1},\tau_j]$ then $\dot{S} = \kappa(d-S)$ integrates exactly to $S(\tau_j) = d + \big(S(\tau_{j-1}) - d\big)e^{-\kappa\Delta\tau}$, which is (11) with $\gamma_j = 1 - e^{-\kappa\Delta\tau}$. Splitting the interval multiplies the retained fraction by $e^{-\kappa\Delta_1}e^{-\kappa\Delta_2} = e^{-\kappa\Delta}$, so inserting an extra measurement of the same d between two probe times leaves S unchanged: the accumulated evidence depends on *where* the probe times are and not on how finely the interval is gridded. Initialised at the first measurement, S is a convex combination of the measurements at every step, by induction on (11). At $\kappa = 4$ and the default spacing $\tau \in \{0.3, 0.6, 0.9\}$ every $\gamma_j = 1 - e^{-1.2} = 0.70$, and the weights on $(d_{0.3}, d_{0.6}, d_{0.9})$ are $(0.091, 0.211, 0.699)$.

Accumulating the *energy* $d = r^2/2$ rather than the magnitude $r = \sqrt{2d}$ is not cosmetic: for weights $w_j \ge 0$ summing to one,

$$\tfrac{1}{2}\sum_j w_j r_j^2 \;=\; \tfrac{1}{2}\Big(\sum_j w_j r_j\Big)^2 \;+\; \tfrac{1}{2}\sum_j w_j\Big(r_j - \sum_l w_l r_l\Big)^2, \tag{38}$$

so filtering the energy measures the persistent mismatch *plus* its variation over time, while filtering the magnitude and squaring afterwards drops the second term. A pixel whose residual is steady and a pixel whose residual flickers with the same mean are then distinguished, and the flickering one – the model disagreeing with itself about what belongs there – is scored higher.

A.3.7 THE MAGNITUDE PARAMETERISATION, AND THE DEFAULT CONSTANTS

The belief (8) is specified in the implementation not by (λ, T) but by a threshold λ_r and a softness T_r on the residual *magnitude* $r = \sqrt{2d}$, which is the velocity-mismatch norm of Section A.3.5 and is the quantity with an interpretable scale. The two agree when

$$\lambda \;=\; \tfrac{1}{2}\lambda_r^2, \qquad T \;=\; \lambda_r T_r. \tag{39}$$

The first matches the hard thresholds exactly, since $d > \lambda \iff r > \lambda_r$. The second matches the slope at the threshold: as a function of r the energy logit is $\big(d(r) - \lambda\big)/T = (r^2 - \lambda_r^2)/(2T)$, whose derivative at $r = \lambda_r$ is λ_r/T, against $1/T_r$ for the magnitude logit. Beyond the threshold the energy logit grows quadratically in r, so large residuals are rejected more decisively than with the magnitude logit.

The defaults are $\lambda_r = 0.4$ and $T_r = 0.05$, hence $\lambda = 0.08$ and $T = 0.02$; the coupling is $\mu = 0.08$, i.e. $\mu/T = 4$ in logit units; the filter rate is $\kappa = 4$; and a run uses $K_p = 32$ probe paths, $K_c = 4$ controlled paths and $C = 2$ cycles at probe times $\tau \in \{0.3, 0.6, 0.9\}$.

A.4 THE CONTROLLER

A.4.1 THE ENDPOINT IDENTITIES

Why $\hat{L}$ and $\hat{U}$ are the endpoints. On the training interpolant $x_t = (1-t)U + tX$ the target velocity is $X - U$, so $x_t + (1-t)(X-U) = (1-t)U + tX + (1-t)X - (1-t)U = X$ and $x_t - t(X-U) = (1-t)U + tX - tX + tU = U$. Replacing $X - U$ by the model's estimate $v_\theta(x,t) \approx \mathbb{E}[X - U \mid x_t = x, t]$ gives the two predictions of (18); for the exact velocity they are the conditional means $\hat{L}_t(x) = \mathbb{E}[X \mid x_t = x]$ and $\hat{U}_t(x) = \mathbb{E}[U \mid x_t = x]$.

The two algebraic identities. With $\hat{L} = x + (1-t)v$ and $\hat{U} = x - tv$,

$$(1-t)\hat{U} + t\hat{L} = x - t(1-t)v + t(1-t)v = x, \tag{40}$$

$$(1-t-\Delta)\hat{U} + (t+\Delta)\hat{L} = x + v\big[-t(1-t-\Delta) + (t+\Delta)(1-t)\big] = x + \Delta v. \tag{41}$$

So recomposing at the current time returns the state unchanged, and recomposing at the next grid time with the *unedited* endpoints is exactly an Euler step. The recursion (19) is therefore an Euler step plus whatever the edit added, which is the content of Section A.4.6.

$\hat{U}$ is conserved along a free path and not along a controlled one. If the model is exact on the interpolant then along an uncontrolled trajectory from U we have $v = X - U$ and $x_t = (1-t)U + tX$, so $\hat{U}_t(x_t) = x_t - tv = U$ at every t: the source estimate is a conserved quantity of the free flow. An edit moves the state off that trajectory, and $\hat{U}$ is then re-estimated at the new state by (18), so it need not equal U and in general does not. This is what Section 3.3 means by the source endpoint entering the recomposition as the model *currently* estimates it: within a step the edit touches only the data endpoint, but across steps neither endpoint is carried.

A.4.2 SOLVING THE CONTROL PROBLEM BY PONTRYAGIN'S PRINCIPLE

This is (13), which Section 3.3 states without proof. Write the terminal cost of (12) as $\Psi(x) = \sum_i (1 - z_i)\|M_i - x_i\|^2/(2s^2)$.

The Hamiltonian and the costate. With the dynamics $\dot{x} = v_\theta(x,t) + a_t$ and running cost $\frac{1}{2}\|a\|^2$,

$$H(x, p, a, t) \;=\; \tfrac{1}{2}\|a\|^2 + p^\top \big(v_\theta(x,t) + a\big),$$

which is strictly convex in a and minimised at

$$a_t \;=\; -\,p_t. \tag{42}$$

Pontryagin's principle gives the costate equation and terminal condition

$$\dot{p}_t \;=\; -\frac{\partial H}{\partial x} \;=\; -J_\theta(x_t, t)^\top p_t, \qquad p_1 \;=\; \nabla\Psi(X_1) \;=\; -\,\frac{(1-z)\odot(M - X_1)}{s^2}, \tag{43}$$

where $J_\theta = \partial v_\theta/\partial x$ is the Jacobian of the velocity field.

Integrating the costate gives the endpoint Jacobian. The costate equation is linear, and its transition matrix is the transpose inverse of the state variational equation's. Concretely, let δ_t solve the variational equation $\dot{\delta}_t = J_\theta(x_t, t)\delta_t$ along the optimal trajectory. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(p_t^\top \delta_t\big) = \dot{p}_t^\top \delta_t + p_t^\top \dot{\delta}_t = -p_t^\top J_\theta \delta_t + p_t^\top J_\theta \delta_t = 0,$$

so $p_t^\top \delta_t$ is constant in t. Since $\delta_1 = (\partial X_1/\partial x_t)\,\delta_t$ for every δ_t, we get $p_t^\top \delta_t = p_1^\top (\partial X_1/\partial x_t)\delta_t$ for all δ_t, hence $p_t = (\partial X_1/\partial x_t)^\top p_1$. Substituting into (42),

$$a_t \;=\; \Big(\frac{\partial X_1}{\partial x_t}\Big)^\top \frac{(1-z)\odot\big(M - X_1\big)}{s^2}, \tag{44}$$

which is (13). □

What it says. Only the trusted pixels drive the control: the vector being propagated is supported on $\{z_i = 0\}$, so nothing in (44) pulls an anomalous pixel towards M by a term of its own. At $t = 1$ the Jacobian is the identity and the control on an anomalous pixel is exactly zero. Earlier it need not be, because $(\partial X_1/\partial x_t)^\top$ mixes pixels: an anomalous pixel is steered only insofar as the flow couples it to the repair of its neighbours.

With the belief optimal instead of known. Minimising (6) over q first leaves the control problem $\min_a \int_0^1 \frac{1}{2}\|a_t\|^2\,\mathrm{d}t + \Phi_\mu\big(d(X_1)\big)$, with Φ_μ the value (26) of the belief problem. By (31) its terminal cost has gradient $-\big(\mathbf{1} - q^\star(d(X_1))\big)\odot(M - X_1)/s^2$, so the same derivation gives (44) with $\mathbf{1} - q^\star\big(d(X_1)\big)$ in place of $1 - z$. This is the stationarity condition stated in Section 3.3. It is necessary, not sufficient, and it is a fixed point rather than a closed form: the belief is evaluated at the endpoint that the control produces.

A.4.3 The endpoint Jacobian, and the step to a single action

Equation (44) is implicit – X_1 is the state the control produces – and for a general velocity field evaluating $\partial X_1/\partial x_t$ means an adjoint pass along the whole trajectory. If the endpoint prediction is taken for the terminal state, $X_1 = \hat{L}_t(x_t) = x_t + s_t v_\theta(x_t, t)$, differentiating gives

$$\frac{\partial X_1}{\partial x_t} \;=\; I + s_t\, J_\theta(x_t, t), \tag{45}$$

and (44) collapses to one vector–Jacobian product at the current state,

$$a_t \;=\; \frac{1}{s^2}\Big[\,(1-z)\odot(M - \hat{L}_t) \;+\; s_t\, J_\theta^\top\big[(1-z)\odot(M - \hat{L}_t)\big]\Big]. \tag{46}$$

Equation (45) is a sanity check as much as a formula: it is I at $t = 1$, where $\hat{L}_1(x) = x$; and it is exactly 0 when the data distribution is a point mass, where $\hat{L}_t$ is constant and no control can move the endpoint.

The single-action problem (15) drops the second term, that is, it takes

$$\frac{\partial X_1}{\partial x_t} \;\approx\; I, \tag{47}$$

and it makes three further replacements, all stated in Section 3.3: the terminal state by the prediction $\hat{L}_t$, the whole control by one action held over $[t, 1]$, and the sensor scale s by the scale s_t of the prediction. The error of (47) is $s_t J_\theta$, which vanishes at $t = 1$ and grows backwards along the path: the surrogate is exact where the controller commits hardest and worst where it commits least. Under (47) the control on $\{z_i = 1\}$ is identically zero at every time, not only at $t = 1$, so the approximation also removes the one mechanism by which an anomalous pixel could be steered at all.

A.4.4 The single-action problem, in full

This is the algebra behind (16) and (17). Everything is per pixel; fix one and drop the index.

The stationarity condition. With $\tilde{L}_{1,t} = \hat{L}_t + s_t a$, the objective (15) is

$$\mathcal{J}(a) \;=\; \frac{s_t^2}{2}\|a\|^2 \;+\; (1-q_t)\,\frac{\left\|M - \hat{L}_t - s_t a\right\|^2}{2s_t^2},$$

strictly convex in a, so its stationary point is the minimum:

$$\frac{\mathrm{d}\mathcal{J}}{\mathrm{d}a} = s_t^2\, a - (1-q_t)\,\frac{M - \hat{L}_t - s_t a}{s_t} = 0.$$

Multiplying by s_t and collecting a,

$$a\big[s_t^3 + s_t(1-q_t)\big] = (1-q_t)\big(M - \hat{L}_t\big) \quad\Longrightarrow\quad a^\star = \frac{(1-q_t)\big(M - \hat{L}_t\big)}{s_t\big(s_t^2 + 1 - q_t\big)},$$

which is (16). The displacement it buys is $s_t a^\star = (1-q_t)(M-\hat{L}_t)/(s_t^2+1-q_t)$, giving the second half of (16).

Why it is a blend and not a projection. The control appears on both sides of $\mathcal{J}$: it displaces the very endpoint it is trying to place. Resolving that implicitness is what produces the denominator. Writing $w = (1-q_t)/s_t^2$ and dividing through,

$$\tilde{L}_{1,t} = \hat{L}_t + \frac{w}{1+w}\big(M - \hat{L}_t\big) = \frac{\hat{L}_t + wM}{1+w}, \qquad M - \tilde{L}_{1,t} = \frac{M - \hat{L}_t}{1+w}, \tag{48}$$

which is (17). Two consequences hold for *every* $w \geq 0$, and therefore for the last step (20) as well, whose weight is $w = (1-q)/q$. First, $\tilde{L}$ is a convex combination of $\hat{L}$ and M, so the step never passes the observation. Second, the residual is the prediction's residual shrunk by $1+w$ and never changes sign: the controller can only reduce disagreement with M, never create it or reverse it.

The pull, at the default schedule. The fraction of the way to M that a pixel is moved is $g = w/(1+w) = (1-q_t)/(s_t^2+1-q_t)$, which is 0 at $q_t = 1$ and $1/(1+s_t^2)$ at $q_t = 0$. At the probe times $\tau \in \{0.3, 0.6, 0.9\}$ that ceiling is 0.67, 0.86 and 0.99: a fully trusted pixel is moved most of the way to the observation at every intermediate step, and an anomalous pixel is not moved at all. The identity (48) is why the readout can be taken as a plain residual against M: the controller's only effect on a pixel is a shrinkage of that residual.

A.4.5 The last step, and why the order matters

Section 3.3 sends $w \to \infty$ at the last step and lands on (20). The limit has to be taken at a *known* support and averaged afterwards; taken in the other order it returns M at every pixel that is not fully rejected, which would copy the anomaly into the restoration. This subsection does both and shows where they part. The support z is binary and is never taken anywhere: the minimiser is computed exactly at each of its two values and the two results are then combined using $\mathbb{P}(z=1) = q$. The only limit is in the scale s, and it is taken because at $t = 1$ the state carries no noise of its own.

(a) The minimiser at a known z. Replacing q_t by z in (48),

$$\tilde{L}(z) = \frac{\hat{L} + w(z)M}{1 + w(z)}, \qquad w(z) = \frac{1-z}{s^2}. \tag{49}$$

(b) Its two values. At $z = 1$, $w = 0$ and $\tilde{L}(1) = \hat{L}$: an anomalous pixel is not moved at all, at any scale. At $z = 0$, $w = 1/s^2$, and multiplying numerator and denominator by s^2,

$$\tilde{L}(0) \;=\; \frac{s^2\hat{L} + M}{s^2 + 1}.$$

(c) The scale goes to zero. At $t = 1$ the state carries no noise, and the only scale left in (1) is the sensor noise, negligible beside a defect. Letting $s \to 0$ in (b) gives $\tilde{L}(0) \to M$ while $\tilde{L}(1) = \hat{L}$ for every s, that is,

$$\lim_{s\to 0} \tilde{L}(z) \;=\; (1-z)\,M + z\,\hat{L}. \tag{50}$$

(d) Averaging is exact here, and only here. Equation (50) is *affine* in z, and an affine function commutes with expectation. With $\mathbb{P}(z=1)=q$, so $\mathbb{E}[z]=q$,

$$\mathbb{E}_z\big[(1-z)M + z\hat{L}\big] = \big(1-\mathbb{E}[z]\big)M + \mathbb{E}[z]\,\hat{L} = (1-q)\,M + q\,\hat{L},$$

which is (20). No approximation is made: substituting q into (50) gives the same number.

(e) The other order does not. At any $s>0$, (49) is a *ratio* in z, not affine, so the two operations do not commute. Substituting q first gives $w=(1-q)/s^2$, which for fixed $q<1$ diverges as $s\to 0$, so

$$\frac{\hat{L}+wM}{1+w} = \frac{\hat{L}/w+M}{1/w+1} \;\longrightarrow\; \frac{0+M}{0+1} = M.$$

Certainty equivalence in the hard limit returns M at every pixel that is not fully rejected. □

Table 3 runs (a)–(e) on one pixel with numbers in it: $\hat{L}=0$, $M=1$, $q=0.9$. The two closures agree to within 0.04 at $s=1$ and disagree by 0.9 by $s=0.01$; the averaged column converges to the last step (20) and the substituted column to the observation.

Table 3: The two orders, on one pixel with $\hat{L}=0$, $M=1$ and $q=0.9$. "Averaged" is $\mathbb{E}_z[\tilde{L}(z)] = (1-q)/(1+s^2)$, the order (20) takes; "substituted" is $\tilde{L}$ evaluated at $q_t=q$, $(1-q)/(s^2+1-q)$. Only the first converges to the posterior mean 0.1.

s	1	0.5	0.2	0.1	0.01
averaged, then $s\to 0$	0.050	0.080	0.096	0.099	0.100
substituted, then $s\to 0$	0.091	0.286	0.714	0.909	0.999

The two orders as two interior rules. The same two orders give two rules for the intermediate steps. Write $r=M-\hat{L}_t$ and $\delta=s_t a$ for the displacement of the endpoint, and let $J_z(\delta)=\frac{1}{2}\|\delta\|^2+(1-z)\|r-\delta\|^2/(2s_t^2)$ be the one-step cost at a known z, so that (15) is $\mathbb{E}_z J_z$ with $\mathbb{P}(z=1)=q_t$. Then

$$\arg\min_\delta \mathbb{E}_z J_z = \frac{1-q_t}{s_t^2+1-q_t}\,r, \qquad \mathbb{E}_z\Big[\arg\min_\delta J_z\Big] = \frac{1-q_t}{s_t^2+1}\,r. \tag{51}$$

The first is the blend (17); the second averages the two known-support minimisers (49) and is the alternative "average of (49) over z" of Table 7. Their gains differ by

$$\frac{1-q_t}{s_t^2+1-q_t} - \frac{1-q_t}{s_t^2+1} \;=\; \frac{q_t(1-q_t)}{(s_t^2+1-q_t)(s_t^2+1)} \;\geq\; 0,$$

which vanishes for a binary belief and is positive in between: the blend pulls a pixel of uncertain belief harder towards M. As $s_t\to 0$ the second gain tends to $1-q_t$, which is the last step (20), and the first tends to 1 for every $q_t<1$, which is (e). The method uses the first rule at the intermediate steps and the limit of the second at the last step.

A.4.6 The edit is an admissible control

What the implementation executes is the edit of (17), and what it recomposes with is (19). That is the same thing as applying a control, by (41):

$$x_{t+\Delta} = (1-t-\Delta)\,\hat{U}_t + (t+\Delta)\,\tilde{L}_t = x_t + \Delta\big[v_\theta(x_t,t)+a_t\big], \qquad a_t = \frac{t+\Delta}{\Delta}\big(\tilde{L}_t-\hat{L}_t\big), \tag{52}$$

since adding $(t+\Delta)(\tilde{L}_t-\hat{L}_t)$ to (41) gives $x_t+\Delta v+(t+\Delta)(\tilde{L}_t-\hat{L}_t)$. So (19) is an Euler step of the controlled flow $\dot{x}=v_\theta(x,t)+a_t$ of (12): the edit is an admissible control, a bounded additive term on the model's own velocity rather than a projection that abandons its dynamics. The executed control is not the planned action of (16), whose displacement is spread over $[t,1]$; the recursion applies it over one grid interval. The implementation accumulates $\frac{1}{2}\Delta\|a_t\|^2$, the integrand of (12), as the control cost of the path.

At the last step $t=\tau_J$ and $t+\Delta=1$, so $a_{\tau_J}=(1-q)\odot(M-\hat{L}_{1,\tau_J})/s_{\tau_J}$, and feeding it back through the Euler step collapses the source endpoint entirely, since $1-t-\Delta=0$:

$$X_1 = x_{\tau_J} + s_{\tau_J}\big[v_\theta + a_{\tau_J}\big] = \hat{L}_{1,\tau_J} + (1-q)\odot\big(M-\hat{L}_{1,\tau_J}\big) = (1-q)\odot M + q\odot\hat{L}_{1,\tau_J},$$

which is (20) again, reached from the control side.

A.4.7 WHAT THE CONTROL CHANGES, IN A LINEAR-GAUSSIAN MODEL

The ablation "last-step edit only" of Table 7 keeps (20) and removes the edits at the intermediate steps. To see what those edits add, take a model in which every quantity is exact. Flatten the image into coordinates and assume: (G1) $L \sim \mathcal{N}(m, \Sigma)$ with $\Sigma \succ 0$; (G2) v_θ is the exact flow-matching velocity of $x_t = (1-t)U + tL$, with $U \sim \mathcal{N}(0, I)$ independent of L; (G3) the anomaly occupies a known set R of coordinates, which is used as the belief at every probe time, $q = \mathbf{1}_R$, with trusted complement N; (G4) the trusted coordinates carry no noise, $M_N = L_N$, and $M_R = L_R + A_R$.

Proposition 5 (linear-Gaussian completion)**.** *Under (G1)–(G3), a controlled path started at $x_0 = U$ ends at*

$$X_{1,N} = M_N, \qquad X_{1,R} = m_R + K\,(M_N - m_N) + Z\,U,$$

where the matrices K and Z depend on Σ, R and the time grid but not on M. If no intermediate step edits the path, so that only the last step applies (20)*, then $K = 0$. Under (G4) in addition,*

$$M_R - X_{1,R} \;=\; A_R + \xi + (K_\star - K)(L_N - m_N) - Z\,U, \qquad K_\star = \Sigma_{RN}\Sigma_{NN}^{-1}, \tag{53}$$

where $\xi = L_R - m_R - K_\star(L_N - m_N) \sim \mathcal{N}(0, \Sigma_{R|N})$ is the part of L_R that the context L_N does not determine, and the three terms after A_R are independent of each other and do not depend on A_R.

Proof. For the exact velocity, $\hat{L}_t$ and $\hat{U}_t$ are the conditional means of L and U given x_t (Section A.4.1). With $y = x - tm$ and $\Omega_t = t^2\Sigma + s_t^2 I$, the covariance of x_t, Gaussian conditioning gives $\hat{L}_t(x) = m + B_t y$ and $\hat{U}_t(x) = N_t y$, with $B_t = t\Sigma\Omega_t^{-1}$ and $N_t = s_t\Omega_t^{-1}$. At an intermediate step with belief $\mathbf{1}_R$ the blend (17) moves the trusted coordinates by the fraction $1/(1+s_t^2)$ and leaves R alone, $\tilde{L} = \hat{L} + G_t(M - \hat{L})$ with $G_t = \mathrm{diag}(\mathbf{1}_N)/(1+s_t^2)$; at the first step $G = 0$. Writing $y_j = x_{t_j} - t_j m$, the recursion (19) becomes

$$y_{j+1} = F_j y_j + t_{j+1} G_{t_j}(M - m), \qquad F_j = s_{t_{j+1}} N_{t_j} + t_{j+1}\big(I - G_{t_j}\big) B_{t_j},$$

with $y_0 = U$. Hence $y_J = C\,U + D\,(M - m)$ for matrices C, D that do not depend on M, and $D(M-m)$ depends on M only through $(M-m)_N$, because every G_{t_j} is zero on R. The last step (20) with $q = \mathbf{1}_R$ keeps M on N and returns $\hat{L}_{1,\tau_J} = m + B_{\tau_J} y_J$ on R, which is the display with $K = [B_{\tau_J} D]_{RN}$ and $Z = [B_{\tau_J} C]_R$, the rows of R. Without intermediate edits every G_{t_j} is zero, so $D = 0$ and $K = 0$. Under (G4), $M_R - X_{1,R} = A_R + (L_R - m_R) - K(L_N - m_N) - ZU$; splitting $L_R - m_R = K_\star(L_N - m_N) + \xi$, with ξ independent of L_N by Gaussian conditioning, gives (53), and U is independent of L. □

The anomaly passes into the residual unchanged, and the control acts only on the terms that do not depend on it. The context term $(K_\star - K)(L_N - m_N)$ has covariance $(K_\star - K)\Sigma_{NN}(K_\star - K)^\top$. Without intermediate edits $K = 0$: the completion of R ignores the context, and the context term has its largest form, $K_\star \Sigma_{NN} K_\star^\top$, so that together with ξ the residual carries the whole prior variation Σ_{RR} of the normal image on R. The intermediate edits make the completion depend on the observed context through K, and at $K = K_\star$, the conditional mean of L_R given L_N, the context term vanishes. The recursion does not guarantee $K = K_\star$, and the source term ZU is present in both cases, so the proposition explains what the intermediate edits can remove rather than proving that every schedule removes it; the ablation measures the difference on the real model.

A.5 THE CYCLES AND THE READOUT

A.5.1 MERGING THE CYCLES

A cycle's accumulated evidence $S^{(c)}$ is its own: the filter state of (11) is *not* carried across the cycle boundary, and cycle c initialises S at its own first measurement. The earlier cycles enter only through the merge, with the operator of Section A.3.4 and the same temperature,

$$\bar{S} \;=\; -T \log \frac{1}{c} \sum_{c'=1}^{c} e^{-S^{(c')}/T}, \tag{54}$$

and the belief is computed from $\bar{S}$, not from $S^{(c)}$, at *every* probe time of cycle c – not only at its end. A later cycle therefore starts with a belief that already carries the earlier ones, though it starts with no belief at all until it has taken its own first measurement, since $S = \perp$ until then.

Proposition 6 (the merge averages the odds)**.** *Let $\tilde{q}^{(c')} = \sigma\big((S^{(c')} - \lambda)/T\big)$ be the uncoupled belief of the evidence of cycle c' and $\bar{q} = \sigma\big((\bar{S} - \lambda)/T\big)$ that of the merged evidence. Then*

$$\frac{1-\bar{q}}{\bar{q}} \;=\; \frac{1}{c}\sum_{c'=1}^{c} \frac{1-\tilde{q}^{(c')}}{\tilde{q}^{(c')}}, \qquad \min_{c'} S^{(c')} \;\le\; \bar{S} \;\le\; \min_{c'} S^{(c')} + T\log c.$$

For two cycles, $\bar{S} > \lambda$ requires $S^{(1)} > \lambda - T\log 2$ and $S^{(2)} > \lambda - T\log 2$, and a rejection by the first cycle ($S^{(1)} > \lambda$) is withdrawn ($\bar{S} < \lambda$) exactly when $\tilde{q}^{(2)} < \tilde{q}^{(1)}/(4\tilde{q}^{(1)} - 1)$.

Proof. $(1-\bar{q})/\bar{q} = e^{-(\bar{S}-\lambda)/T} = \frac{1}{c}\sum_{c'} e^{-(S^{(c')}-\lambda)/T}$, and each summand is $(1-\tilde{q}^{(c')})/\tilde{q}^{(c')}$. The bounds are Proposition 3 with $K = c$. For two cycles, $\bar{S} \le \min_{c'} S^{(c')} + T\log 2$ gives the first claim. Writing $o_{c'} = (1-\tilde{q}^{(c')})/\tilde{q}^{(c')}$, $\bar{S} < \lambda$ is $o_1 + o_2 > 2$; with $\tilde{q}^{(1)} > \frac{1}{2}$ this is $1/\tilde{q}^{(2)} > (4\tilde{q}^{(1)} - 1)/\tilde{q}^{(1)}$, which is the stated condition. □

The merge is therefore a conservative confirmation. After two cycles a pixel is rejected only if neither cycle's evidence falls below $\lambda - T\log 2$, and a later cycle can withdraw an earlier rejection: a first-cycle belief of 0.9 is withdrawn when the second cycle's falls below 0.346. That is the reading the cycles are designed for. Cycle c probes towards the previous restoration Y_c while still measuring every residual against M, so by (37) a rejected pixel whose previous completion agrees with M loses its evidence in the new cycle, while one whose completion departs from M keeps it in every cycle and the merged evidence stays high. The alternatives in Table 11 are the mean, the maximum, and dropping the earlier cycles. With $\mu > 0$ the coupled belief $\mathcal{G}_\mu(\bar{S})$ can still reject or trust a pixel through its neighbours (Section A.3.2).

A.5.2 The current control never reaches the belief

Proposition 7 (same-cycle invariance)**.** *Suppose the network evaluates each sample of a batch as if it were alone. Then within cycle c the energies, the evidence S and $\bar{S}$, and the belief q at every probe time are functions of M, v_θ, the probe target Y_c, the probe sources of cycle c, the evidence $S^{(1)}, \dots, S^{(c-1)}$ of the finished cycles, the warm start of the belief solver and the fixed settings, and of nothing else. In particular they are the same whatever the controlled paths of cycle c do.*

Proof. The probe states $P_\tau^k = (1-\tau)U^k + \tau Y_c$ are fixed arrays. By the assumption, the rows of the batched network call that belong to the probes do not depend on the other rows. Pooling, the filter, the merge and the belief solve are deterministic functions of those rows and of the listed inputs. □

The U-Net used here normalises with group normalisation inside each sample and shares no statistic across the batch, so the assumption holds up to floating-point rounding. This is the separation that Section 3.1 asks for, and it is causal rather than statistical: path k and probe k share the source U^k (Algorithm 1), which correlates them without letting the controlled state act on the probe. Earlier control does reach later cycles, through the probe target Y_c (37) from the second cycle on, and through the stored evidence and the warm start from the third.

A.5.3 The last-step identity and the readout

From (20), subtracting from M,

$$X_1 = (1 - q_{\tau_J}) \odot M + q_{\tau_J} \odot \hat{L}_{1,\tau_J}, \qquad M - X_1 = q_{\tau_J} \odot \big(M - \hat{L}_{1,\tau_J}\big), \tag{55}$$

which is (48) at $w = (1-q)/q$. The residual of a restoration is the belief *times* the residual of the model's own completion, so it carries two statements about a pixel – whether the observer rejects it, and by how much the completion disagrees with the observation – and a pixel needs both to score highly. The readout needs no separate estimator: it is the restoration's residual, which already carries the belief.

The readout (21) takes the maximum over the K_c controlled paths. Since the belief q_{τ_J} is shared by the paths and nonnegative, it factors out of the maximum exactly:

$$A_i \;=\; q_{\tau_J,i} \max_{k=1,\dots,K_c} \big\|M_i - \hat{L}^k_{1,\tau_J,i}\big\|. \tag{56}$$

Table 4: Reference configuration of FlowCTRL, in the notation of Section 3. Each group header names the table in which the group is varied. All runs use bfloat16 arithmetic and a compiled network; running in float32 or without compilation changes each metric by less than 0.08 points on MVTec-AD and less than 0.15 points on VisA. The same settings are used on both datasets.

Setting	Value
Belief (Section 3.2); varied in Table 5	
Threshold λ	0.080
Temperature T	0.020
Coupling μ/λ	1
FISTA iterations for (7)	100
Controller (Section 3.3); varied in Tables 7 and 8	
Controlled paths K_c	4
Time grid	probe times, $J + 1 = 4$ steps per cycle
Edit at the intermediate steps	blend (17)
Edit at the last step	posterior mean (20)
Steering window	all steps
Observer (Section 3.2); varied in Tables 9 and 10	
Probe paths K_p	32
Probe times $\tau_1, \ldots, \tau_J$	$0.3, 0.6, 0.9$
Energy measured on	probe paths (9)
Pooling over sources	soft minimum (10)
Accumulation over probe times	filter (11)
Filter rate κ	4
Accumulated quantity	energy d
Cycles and readout (Section 3.4); varied in Table 11	
Cycles C	2
Probe target of cycles $c > 1$	previous restoration
Sources of cycles $c > 1$	new draws
Merge of the cycle evidence	soft minimum (54)
Pooling of the K_c residuals	maximum (21)
Reported map	residual (21)
Post-processing (Section 3.4); varied in Table 12	
Gaussian smoothing	kernel 9×9, $\sigma = 3$
Image score	mean of the top 1% of A

The map is the belief times the largest completion residual. On a pixel the belief trusts, every residual is close to zero, and so is the map. On a pixel it rejects, the map is the largest distance between the observation and a path's completion, so a pixel counts as anomalous as soon as *one* completion departs from it. The soft minimum of Section A.3.4, which would call a pixel normal as soon as one completion reproduces it, is one of the alternatives in Table 11; with it the belief no longer factors out at the same temperature.

APPENDIX B REFERENCE CONFIGURATION AND ABLATION RESULTS

B.1 REFERENCE CONFIGURATION

Table 4 lists the settings of the reference run in the notation of Section 3. The settings were chosen by sweeping on the test split of MVTec-AD; Section B.3 reports one-at-a-time sweeps around these values. The implementation sets the threshold λ and the temperature T through a threshold and a temperature on the magnitude $\sqrt{2d}$; Section A.3.7 gives the conversion.

B.2 ABLATION SETTINGS

Protocol. Each part of the method has numeric settings and alternative forms. We sweep each numeric setting over values around its reference value, and we replace single steps of the method by the alternative forms described below. Unless stated otherwise, every run changes one setting of the reference configuration (Table 4) and keeps all other settings fixed. Every ablation run is made

on MVTec-AD and on VisA with seed 0, the seed of the reference run; this reference differs from the 16-seed mean in Table 1 by at most 0.13 points on every metric of both datasets. The rows of Table 2, the points of Figure 3 and the threshold, temperature and coupling runs of the belief are also run with seeds 1 and 2 (Section B.4). Besides the seven metrics, the tables of Section B.3 evaluate the final belief at $q > \frac{1}{2}$ against the ground truth, with three columns defined there.

Budget. The cost of a run grows with the numbers of probe paths K_p, controlled paths K_c, probe times J and cycles C (Equation (22)). Figure 3 shows image AUROC, pixel AP and AUPRO against the time per image when K_p, J or C is varied and the other settings are kept at the reference, as the mean over seeds 0 to 2. On MVTec-AD, fewer probe paths, probe times or cycles than in the reference lower all three metrics; with $J = 1$, pixel AP is 51.4. More probe paths or cycles change each metric by at most 0.5 points, and more probe times lower pixel AP, to 65.8 at $J = 9$, while the time per image increases by up to $2.8\times$. On VisA, fewer probe paths, probe times or cycles also lower all three metrics, and more probe paths or cycles change each metric by at most 0.4 points; with five or nine probe times, pixel AP is 41.8 in both cases, against 40.3 at $J = 3$. Varying K_c from 1 to 16 raises pixel AP from 67.0 to 68.4 and pixel F_1-max from 64.4 to 65.4 on MVTec-AD, at 0.26 to 0.37 s per image, and changes the other metrics by at most 0.2 points; on VisA it lowers AUPRO from 90.7 to 89.6 and changes every other metric by at most 0.4 points (Table 8).

Belief. The belief is computed from (7) with four numeric settings (Table 5): the threshold $\lambda \in \{0.045, 0.080, 0.125, 0.180\}$, the temperature $T \in \{0.010, 0.020, 0.030, 0.040\}$, the coupling $\mu/\lambda \in \{0, 0.25, 0.5, 1, 2\}$, and the number of FISTA iterations of the dual solver (Section A.3.2), $\{10, 25, 50, 100\}$. The runs with $\lambda = 0.045$ and $\lambda = 0.125$ also change T, to 0.015 and 0.025, because they change the threshold on the magnitude $\sqrt{2d}$ at a fixed temperature on that magnitude (Section A.3.7). With $\mu = 0$, (7) reduces to the per-pixel belief (8). Table 6 crosses the four thresholds with $\mu/\lambda \in \{0.25, 0.5, 1, 2\}$. Removing the coupling lowers pixel AP from 67.9 to 61.3 on MVTec-AD and from 40.3 to 37.9 on VisA (Table 2). On MVTec-AD, the reference has the highest image AUROC of the 16 combinations in Table 6, and $\mu/\lambda = 2$ gives a pixel AP higher by 0.2 points and an AUPRO lower by 1.7 points. On VisA, $\mu/\lambda = 2$ gives the highest image AUROC and pixel AP of the 16 combinations, 97.3 and 41.6 against 96.8 and 40.3 at the reference, and an AUPRO lower by 0.8 points.

Controller. The controller has three alternative forms (Table 7). (i) No control: the belief is computed but every step is an unedited flow step, so X_1 is the endpoint of an uncontrolled sample; this run uses one cycle and is compared with the one-cycle run with control. (ii) At the intermediate steps, the blend (17) is replaced by the average of the known-support minimiser (49) over z with $\mathbb{P}(z = 1) = q$, by the blend with the weight $w = (1 - q_t)/s_t$ in place of $(1 - q_t)/s_t^2$, or by the posterior mean (20). (iii) At the last step, the posterior mean (20) is replaced by the blend (17). The controller has four numeric settings (Table 8). A gain $\eta \in \{10^{-6}, 0.1, 0.25, 0.5, 1, 2, 4\}$ multiplies the weight w of (17); the last step applies (20) for every $\eta > 0$, so at $\eta = 10^{-6}$ only the last step edits the path. The number of controlled paths is $K_c \in \{1, 2, 4, 8, 16\}$. The steering window restricts the edits to $t \geq 0.6$ or to $t \leq 0.8$; a step outside the window is an unedited flow step. The time grid is either the probe times or a uniform grid of $N \in \{10, 20, 40\}$ steps; on the uniform grid, the belief is updated only at the steps that fall on a probe time. Without control, pixel AP is 20.4 on MVTec-AD and 7.5 on VisA, against 66.5 and 39.9 with control at $C = 1$ (Table 2); with the last-step edit alone ($\eta = 10^{-6}$, $C = 1$), it is 60.8 and 30.0. Pixel AP is highest at $\eta = 1$ on both datasets; $\eta = 2$ and $\eta = 4$ raise image AUROC by 0.1 points and lower pixel AP. The posterior mean at every step raises pixel AP to 68.3 on MVTec-AD and lowers it to 38.9 on VisA.

Observer. The observer has five alternative forms (Table 9). (i) The energy is measured on the controlled path at every step of the time grid, instead of on the probe paths at the probe times; the belief then exists from the first step, so the edits start at $t = 0$. (ii) The K_p energies are pooled by their mean instead of the soft minimum (10), or each source keeps its own evidence and belief, and the K_p beliefs are averaged or their logits are summed before the sigmoid. (iii) The K_p sources are drawn as antithetic pairs $(U, -U)$ instead of independently. (iv) The measurements are accumulated over the probe times by a weighted mean with weights τ_j^2, or only the last measurement is kept, instead of the filter (11). (v) The magnitude $\sqrt{2d}$ is accumulated or thresholded instead of the energy d. The implementation measures the energy on the controlled path only on a uniform grid,

so this form is compared with the probe paths on the uniform grid with $N = 10$. The coupling in (7) acts on one pooled energy map, so the two per-source poolings are run with $\mu = 0$ and compared with the soft minimum at $\mu = 0$. The numeric settings (Table 10) are the number of probe paths $K_p \in \{4, 8, 16, 32, 64, 128\}$, the filter rate $\kappa \in \{1, 2, 4, 8, 16\}$, and the probe times: we use 1, 2, 3, 4, 5 or 9 probe times ending at 0.9, and five placements of three probe times. Measuring the energy on the controlled path lowers pixel AP from 65.9 to 27.3 on MVTec-AD and from 39.6 to 8.9 on VisA, and pooling by the mean energy lowers it from 67.9 to 33.2 and from 40.3 to 15.5 (Table 2). Antithetic sources change every metric by at most 0.2 points on both datasets.

Cycles and readout. The cycles have three alternative forms (Table 11). (i) The evidence of the cycles is merged by the mean, by the maximum or by keeping the last cycle only, instead of the soft minimum (54). (ii) The cycles after the first probe towards M instead of towards the previous restoration, with new or with reused sources. (iii) The cycles after the first reuse the sources of the first cycle. The numeric setting is the number of cycles $C \in \{1, 2, 3, 4\}$. For the readout, the K_c residuals in (21) are pooled by their soft minimum or their mean energy, or the energy of the mean restoration is used, instead of the maximum. The run also produces the merged probe evidence $\sqrt{2\bar{S}}$ (54) and the belief q as maps, and Table 11 scores them with the same post-processing. All other results use the residual (21), which is the output of the controlled paths. One cycle gives 66.5 pixel AP at 0.15 s per image, and two cycles give 67.9 at 0.28 s; on VisA, one cycle gives 39.9 at 0.15 s and two give 40.3 at 0.29 s (Table 2). On MVTec-AD, the maximum gives a higher image AUROC and pixel AP than the three other poolings of the residuals and an AUPRO lower by at most 0.2 points; on VisA, the three other poolings give an image AUROC higher by 0.1 points, a pixel AP higher by 0.2 to 0.3 points and an AUPRO higher by 0.5 to 1.2 points.

Post-processing. Before scoring, the map is smoothed by a $k \times k$ Gaussian kernel with standard deviation σ ($k = 9$ and $\sigma = 3$ in the reference), and the image score is the mean of the top 1% of the smoothed map. We vary $k \in \{1, 5, 7, 9, 13, 17\}$ at $\sigma = 3$, where $k = 1$ means no smoothing, and $\sigma \in \{1, 2, 3, 4.5, 6\}$ at $k = 9$. For the image score, we use the mean of the top 0.5%, 2% or 5% of the map, its maximum, or the average of the maximum and the top-1% mean. These settings are applied to the saved unsmoothed maps of the reference run, so the controller is not run again (Table 12). Over the smoothing settings, image AUROC changes by at most 0.7 points on MVTec-AD and 0.1 points on VisA. Pixel AP is highest at $\sigma = 1$ on both datasets, 69.2 and 43.7 against 67.8 and 40.3 at the reference, and lowest without smoothing on MVTec-AD (64.5) and at $\sigma = 6$ on VisA (39.0). The top-1% mean gives the highest image AUROC of the six image scores on both datasets.

Table 5: Ablation of the belief (Section 3.2) on MVTec-AD (top) and VisA (bottom), seed 0, in %. The belief columns are defined in Section B.3. The runs with $\lambda = 0.045$ and $\lambda = 0.125$ change the threshold on the magnitude $\sqrt{2d}$ at a fixed temperature on that magnitude (Section A.3.7), which also changes T; its value is given in parentheses.

	Image-level			Pixel-level				Belief $q > \frac{1}{2}$			
Setting	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO	normal	recall	FRR	s/img
MVTec-AD											
Threshold λ											
0.045 ($T = 0.015$)	95.0	97.7	94.8	97.9	65.0	62.5	92.0	22.4	93.1	29.2	0.28
0.080 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
0.125 ($T = 0.025$)	97.3	99.0	96.5	97.9	66.1	63.8	92.1	0.1	52.6	0.6	0.28
0.180	95.4	98.4	95.6	97.6	60.9	59.0	92.1	0.0	33.5	0.2	0.28
Temperature T											
0.010	97.5	99.0	96.2	98.0	66.7	63.8	93.1	0.2	59.0	0.8	0.28
0.020 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
0.030	97.2	98.8	96.2	97.9	67.0	64.3	92.6	5.4	84.8	7.3	0.27
0.040	96.5	98.5	95.6	97.6	65.9	63.4	91.9	15.1	90.7	20.5	0.28
Coupling μ/λ											
0	94.8	97.9	94.6	96.7	61.3	59.1	91.0	7.5	64.9	8.6	0.23
0.25	96.1	98.4	95.5	97.5	64.2	61.4	92.6	3.3	70.1	4.1	0.29
0.5	97.0	98.7	95.9	97.9	66.1	63.1	93.2	2.9	73.1	3.4	0.28
1 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
2	97.3	99.0	96.6	98.0	68.0	65.9	91.3	2.9	72.8	3.1	0.28
FISTA iterations of the dual solver											
10	97.5	98.9	96.4	97.9	67.1	64.2	93.5	3.1	78.6	4.0	0.24
25	97.6	99.0	96.6	98.1	67.7	64.9	93.3	2.8	75.5	3.3	0.25
50	97.6	99.0	96.6	98.1	67.8	65.0	93.2	2.9	75.1	3.2	0.26
100 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
VisA											
Threshold λ											
0.045 ($T = 0.015$)	93.4	94.4	90.9	98.3	36.5	42.8	87.6	29.5	98.0	33.2	0.29
0.080 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
0.125 ($T = 0.025$)	97.2	97.7	94.1	98.7	40.9	46.5	90.8	0.1	67.4	0.8	0.27
0.180	96.5	97.4	93.2	98.4	37.0	43.7	91.8	0.0	49.8	0.3	0.28
Temperature T											
0.010	96.7	97.3	93.9	98.6	38.6	45.1	91.6	0.2	72.3	1.2	0.28
0.020 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
0.030	96.1	96.8	93.4	98.5	40.6	46.5	89.3	9.6	93.5	14.8	0.28
0.040	95.2	96.3	92.7	98.3	40.6	46.5	88.9	21.8	96.0	25.2	0.28
Coupling μ/λ											
0	92.8	94.2	89.9	97.6	38.0	44.5	87.9	10.5	79.4	12.5	0.24
0.25	94.7	95.5	92.1	98.2	38.6	45.0	89.5	4.2	83.2	6.6	0.27
0.5	95.8	96.4	93.4	98.5	39.2	45.4	90.0	2.5	85.0	5.2	0.28
1 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
2	97.3	97.6	94.5	98.7	41.6	47.6	89.3	0.5	84.5	3.3	0.29
FISTA iterations of the dual solver											
10	96.5	97.0	93.8	98.5	39.7	45.8	90.5	3.2	88.0	6.2	0.25
25	96.8	97.2	94.1	98.7	40.1	46.0	90.1	1.8	86.5	4.7	0.25
50	96.8	97.2	94.3	98.7	40.3	46.2	90.0	1.4	86.3	4.2	0.26
100 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29

B.3 Ablation tables

Tables 5 to 12 list all ablation runs of Section B.2, grouped by the part of the method. Each table holds both datasets, MVTec-AD in the upper half and VisA in the lower half with the same blocks and rows; Tables 6 and 12 put the two datasets side by side. Each block varies one setting of Table 4; the row marked "(default)" is the reference run and is repeated in each block. All runs use seed 0, and the metrics are means over the categories, in %. The three belief columns evaluate the final belief at $q > \frac{1}{2}$ against the ground truth. *normal* is the share of the pixels of a normal image with $q > \frac{1}{2}$, averaged over the normal images. *recall* is the share of the defect pixels with $q > \frac{1}{2}$, and FRR, the false-rejection rate, is the share of the non-defect pixels with $q > \frac{1}{2}$, over the normal images and the normal part of the anomalous images; both pool the pixels of all images of a category. Each of the three is computed per category and averaged over the categories. s/img is the controller time per image (Section 4.1).

Table 6: Threshold λ against coupling μ/λ of the belief on MVTec-AD and VisA, seed 0, in %. Every cell is one run of the reference configuration with λ and μ/λ changed; as in Table 5, the runs with $\lambda = 0.045$ and $\lambda = 0.125$ also change T, to 0.015 and 0.025. The underlined cell is the reference. The row $\lambda = 0.080$ and the column $\mu/\lambda = 1$ are runs of Table 5.

	MVTec-AD, μ/λ				VisA, μ/λ			
	0.25	0.5	1	2	0.25	0.5	1	2
Image AUROC								
$\lambda = 0.045$	92.5	93.4	95.0	96.9	89.8	91.5	93.4	95.2
$\lambda = 0.080$	96.1	97.0	97.7	97.3	94.7	95.8	96.8	97.3
$\lambda = 0.125$	97.3	97.6	97.3	96.1	96.5	96.9	97.2	97.0
$\lambda = 0.180$	96.7	96.1	95.4	94.3	96.9	96.9	96.5	96.1
Pixel AP								
$\lambda = 0.045$	60.4	62.4	65.0	65.9	34.9	35.5	36.5	37.6
$\lambda = 0.080$	64.2	66.1	67.8	68.0	38.6	39.2	40.3	41.6
$\lambda = 0.125$	64.2	65.3	66.1	64.5	39.0	39.7	40.9	41.5
$\lambda = 0.180$	60.5	60.7	60.9	60.3	36.9	36.9	37.0	36.0
AUPRO								
$\lambda = 0.045$	90.4	91.3	92.0	91.4	86.4	87.0	87.6	87.6
$\lambda = 0.080$	92.6	93.2	93.0	91.3	89.5	90.0	90.1	89.3
$\lambda = 0.125$	92.9	93.1	92.1	89.7	91.1	91.4	90.8	89.1
$\lambda = 0.180$	91.9	92.3	92.1	91.5	92.0	92.1	91.8	91.6
FRR of the belief								
$\lambda = 0.045$	24.6	27.3	29.2	30.5	31.3	32.2	33.2	34.3
$\lambda = 0.080$	4.1	3.4	3.3	3.1	6.6	5.2	4.2	3.3
$\lambda = 0.125$	1.2	0.8	0.6	0.5	1.6	1.1	0.8	0.6
$\lambda = 0.180$	0.4	0.3	0.2	0.2	0.5	0.4	0.3	0.2

Table 7: Ablation of the controller (Section 3.3), alternative forms, on MVTec-AD (top) and VisA (bottom), seed 0, in %. The belief columns are defined in Section B.3. The first block of each half uses one cycle ($C = 1$); at $\eta = 10^{-6}$ only the last step edits the path (Table 8).

	Image-level			Pixel-level				Belief $q > \frac{1}{2}$			
Setting	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO	normal	recall	FRR	s/img
MVTec-AD											
Control ($C = 1$)											
with control	97.6	99.0	96.4	98.0	66.5	64.0	92.8	2.1	71.3	2.6	0.15
last-step edit only, $\eta = 10^{-6}$	93.8	96.9	94.2	96.9	60.8	60.0	87.8	2.1	71.3	2.6	0.15
no control	63.7	82.7	85.2	80.3	20.4	25.3	50.0	2.1	71.3	2.6	0.15
Edit at the intermediate steps											
blend (17) (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
average of (49) over z	96.9	98.6	96.0	98.1	67.5	64.9	92.3	3.5	75.8	3.6	0.28
blend with $w = (1 - q_t)/s_t$	97.1	98.8	96.1	98.0	67.7	65.0	92.3	3.5	75.7	3.6	0.28
posterior mean (20)	97.4	98.9	96.2	98.1	68.3	65.4	93.0	3.2	75.3	3.4	0.28
Edit at the last step											
posterior mean (20) (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
blend (17)	96.8	98.9	96.2	97.3	62.9	60.0	92.0	1.5	70.0	2.0	0.28
VisA											
Control ($C = 1$)											
with control	96.5	97.0	93.8	98.6	40.0	45.7	89.9	1.0	81.6	3.5	0.15
last-step edit only, $\eta = 10^{-6}$	89.8	92.4	87.2	97.4	30.0	37.3	85.1	1.0	81.6	3.5	0.15
no control	63.1	68.7	74.7	86.3	7.5	10.8	56.2	1.0	81.6	3.5	0.15
Edit at the intermediate steps											
blend (17) (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
average of (49) over z	96.0	96.3	93.5	98.7	38.7	44.6	89.1	1.5	87.8	4.5	0.29
blend with $w = (1 - q_t)/s_t$	96.0	96.6	93.2	98.7	39.7	45.9	89.4	1.6	87.8	4.7	0.28
posterior mean (20)	96.1	96.4	93.4	98.7	38.9	45.1	89.4	1.2	86.2	4.1	0.29
Edit at the last step											
posterior mean (20) (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
blend (17)	96.8	97.4	93.7	98.1	37.2	43.7	91.8	0.6	80.8	2.9	0.28

Table 8: Ablation of the controller (Section 3.3), numeric settings, on MVTec-AD (top) and VisA (bottom), seed 0, in %. The belief columns are defined in Section B.3. The gain η multiplies the weight w of (17); the last step applies (20) for every $\eta > 0$, so at $\eta = 10^{-6}$ only the last step edits the path. With the window $t \leq 0.8$, the last step is an unedited flow step.

	Image-level			Pixel-level				Belief $q > \frac{1}{2}$			
Setting	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO	normal	recall	FRR	s/img
MVTec-AD											
Gain η of the blend weight											
10^{-6} (last-step edit only)	91.7	95.9	93.1	96.6	57.2	57.3	86.5	4.0	77.6	4.7	0.27
0.1	94.5	97.3	94.5	97.2	61.8	60.6	89.0	3.9	77.0	4.2	0.27
0.25	96.2	98.3	95.5	97.7	65.3	62.9	90.9	3.7	76.3	3.8	0.29
0.5	97.1	98.8	96.0	98.0	67.6	65.0	92.3	3.4	75.6	3.5	0.28
1 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
2	97.8	99.1	96.6	98.0	67.3	64.8	93.1	2.8	74.8	3.1	0.27
4	97.8	99.1	96.7	97.8	66.7	64.4	92.9	2.7	74.4	3.0	0.28
Steering window											
all steps (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
$t \geq 0.6$	97.2	98.8	96.2	98.0	66.5	64.2	92.1	3.3	75.5	3.4	0.28
$t \leq 0.8$	96.0	98.5	95.1	96.6	63.7	62.4	90.6	3.7	77.0	4.0	0.28
Controlled paths K_c											
1	97.5	99.0	96.4	98.1	67.0	64.4	93.1	3.1	75.1	3.3	0.26
2	97.6	99.0	96.6	98.1	67.5	64.8	93.0	3.1	75.1	3.3	0.27
4 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
8	97.7	99.0	96.5	98.1	68.2	65.3	93.0	3.0	75.1	3.3	0.30
16	97.7	99.0	96.5	98.1	68.4	65.4	92.9	3.0	75.1	3.3	0.37
Time grid											
probe times (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
uniform, $N = 10$	97.8	99.1	96.8	97.6	65.9	63.7	92.7	2.0	72.3	2.4	0.38
uniform, $N = 20$	97.1	98.9	96.6	97.4	63.7	61.9	92.5	1.3	68.4	1.7	0.54
uniform, $N = 40$	96.2	98.6	95.9	97.5	61.8	60.8	92.0	1.2	68.0	1.7	0.90
VisA											
Gain η of the blend weight											
10^{-6} (last-step edit only)	87.3	90.4	85.5	96.8	25.7	32.7	82.4	5.0	89.5	8.3	0.28
0.1	90.8	93.0	88.0	97.7	31.2	38.2	85.4	3.9	89.2	7.3	0.28
0.25	94.0	95.4	91.3	98.3	36.2	43.1	87.8	2.4	88.8	5.9	0.28
0.5	96.0	96.6	93.3	98.6	39.5	45.8	89.4	1.6	87.7	4.7	0.28
1 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
2	96.9	97.3	93.8	98.5	40.1	46.0	90.2	1.2	85.2	3.9	0.28
4	96.9	97.4	93.9	98.0	39.8	45.7	90.3	1.1	84.1	3.8	0.27
Steering window											
all steps (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
$t \geq 0.6$	96.3	97.1	93.9	98.6	39.1	45.6	89.6	1.6	87.8	4.8	0.29
$t \leq 0.8$	94.1	95.8	91.4	97.2	38.9	45.9	87.0	2.2	88.5	5.5	0.29
Controlled paths K_c											
1	96.9	97.1	94.2	98.8	40.2	46.3	90.7	1.3	86.4	4.2	0.26
2	96.8	97.2	94.2	98.7	40.3	46.3	90.4	1.3	86.3	4.2	0.27
4 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
8	96.8	97.2	94.2	98.7	40.3	46.1	89.8	1.3	86.3	4.2	0.31
16	96.7	97.1	94.2	98.6	40.2	46.0	89.6	1.3	86.4	4.1	0.36
Time grid											
probe times (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
uniform, $N = 10$	96.8	97.4	93.6	97.9	39.7	45.4	90.9	0.8	82.0	3.2	0.38
uniform, $N = 20$	96.5	97.4	93.3	98.0	38.8	44.7	91.2	0.4	78.8	2.5	0.55
uniform, $N = 40$	96.1	97.3	92.6	98.1	37.7	43.8	91.2	0.4	78.6	2.5	0.89

Table 9: Ablation of the observer (Section 3.2), alternative forms, on MVTec-AD (top) and VisA (bottom), seed 0, in %. The belief columns are defined in Section B.3. The first block of each half uses the uniform grid with $N = 10$, on which the controlled-path observer measures the energy at every step; the implementation allows this observer only on a uniform grid. The third block uses $\mu = 0$, because the coupling in (7) acts on one pooled energy map.

	Image-level			Pixel-level				Belief $q > \frac{1}{2}$			
Setting	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO	normal	recall	FRR	s/img
MVTec-AD											
Where the energy is measured, uniform grid ($N = 10$)											
probe paths	97.8	99.1	96.8	97.6	65.9	63.7	92.7	2.0	72.3	2.4	0.38
controlled path	71.6	85.4	87.8	88.8	27.4	31.3	67.8	23.5	80.2	23.8	0.33
Pooling over sources											
soft minimum (10) (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
mean energy	76.5	88.5	87.9	90.4	33.1	36.6	74.9	57.6	97.8	57.8	0.28
Pooling over sources, no coupling ($\mu = 0$)											
soft minimum	94.8	97.9	94.6	96.7	61.3	59.1	91.0	7.5	64.9	8.6	0.23
mean of per-source beliefs	86.4	93.6	90.4	94.6	50.9	52.0	85.6	34.2	88.6	35.9	0.24
sum of per-source logits	74.9	87.8	87.2	90.5	33.6	37.5	72.9	42.2	91.8	43.9	0.23
Probe sources											
independent draws (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
antithetic pairs $(U, -U)$	97.8	99.0	96.7	98.1	67.9	65.1	92.9	2.9	74.9	3.2	0.28
Accumulation over probe times											
filter (11) (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
weighted mean, weights τ_j^2	97.7	99.0	96.6	98.1	67.7	64.9	93.1	2.6	74.1	3.0	0.28
last measurement only	95.0	97.8	95.0	97.5	65.2	63.1	89.8	3.9	72.3	4.2	0.28
Accumulated and thresholded quantity											
energy d (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
accumulate $\sqrt{2d}$	97.7	99.0	96.5	98.1	67.8	64.9	92.7	2.0	71.2	2.4	0.28
threshold $\sqrt{2d}$	97.8	99.1	96.8	98.1	67.8	65.3	92.4	1.3	69.9	1.8	0.28
VisA											
Where the energy is measured, uniform grid ($N = 10$)											
probe paths	96.8	97.4	93.6	97.9	39.7	45.4	90.9	0.8	82.0	3.2	0.38
controlled path	65.0	70.3	75.1	89.8	9.1	12.2	61.5	21.5	92.1	24.9	0.33
Pooling over sources											
soft minimum (10) (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
mean energy	72.1	76.8	77.4	92.7	15.5	19.6	71.4	57.7	99.3	58.8	0.28
Pooling over sources, no coupling ($\mu = 0$)											
soft minimum	92.8	94.2	89.9	97.6	38.0	44.5	87.9	10.5	79.4	12.5	0.24
mean of per-source beliefs	83.6	87.4	82.9	96.7	32.8	39.0	83.2	35.0	94.9	37.3	0.23
sum of per-source logits	72.5	77.3	77.6	92.1	15.6	20.4	68.4	44.0	96.4	45.7	0.23
Probe sources											
independent draws (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
antithetic pairs $(U, -U)$	96.9	97.2	94.2	98.7	40.4	46.2	90.1	1.2	86.3	4.1	0.28
Accumulation over probe times											
filter (11) (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
weighted mean, weights τ_j^2	96.7	97.2	94.2	98.7	40.0	45.9	90.1	1.0	85.1	3.7	0.29
last measurement only	96.2	97.1	93.4	98.5	39.4	46.3	89.7	6.7	88.4	9.3	0.28
Accumulated and thresholded quantity											
energy d (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
accumulate $\sqrt{2d}$	96.9	97.2	94.1	98.7	40.5	46.4	90.1	0.8	83.5	3.3	0.28
threshold $\sqrt{2d}$	97.2	97.5	94.3	98.7	41.2	47.0	89.8	0.6	82.4	2.9	0.28

Table 10: Ablation of the observer (Section 3.2), numeric settings, on MVTec-AD (top) and VisA (bottom), seed 0, in %. The belief columns are defined in Section B.3. All probe schedules of the second block end at 0.9.

	Image-level			Pixel-level				Belief $q > \frac{1}{2}$			
Setting	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO	normal	recall	FRR	s/img
MVTec-AD											
Probe paths K_p											
4	90.5	94.9	93.0	97.1	59.6	58.7	90.1	11.9	89.6	16.4	0.15
8	95.3	97.9	95.2	97.9	65.4	62.9	92.3	5.2	82.8	6.5	0.17
16	97.2	98.8	96.1	98.1	67.3	64.5	92.8	4.0	78.1	4.5	0.21
32 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
64	97.9	99.1	96.6	98.1	68.0	65.3	92.8	1.9	72.6	2.3	0.42
128	97.8	99.1	96.7	98.1	67.9	65.4	92.7	1.3	70.6	1.8	0.72
Number of probe times											
0.9	85.1	92.6	91.4	95.3	51.2	51.9	81.7	5.7	75.5	7.0	0.11
0.6, 0.9	97.0	98.7	96.3	97.9	65.9	63.4	91.9	3.2	74.0	3.3	0.20
0.3, 0.6, 0.9 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
0.3, 0.5, 0.7, 0.9	97.8	99.1	96.6	97.9	67.1	64.3	93.2	1.5	71.8	2.3	0.36
0.1, 0.3, . . . , 0.9 (5 times)	97.5	99.0	96.0	98.0	67.3	64.3	93.1	1.8	72.7	2.5	0.43
0.1, 0.2, . . . , 0.9 (9 times)	97.4	99.0	96.3	97.8	65.8	62.7	92.9	0.5	68.0	1.6	0.76
Placement of three probe times											
0.3, 0.6, 0.9 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
0.15, 0.45, 0.9	95.9	98.3	95.5	97.9	66.9	64.5	91.5	3.8	77.0	4.2	0.27
0.2, 0.5, 0.8	97.0	98.8	96.1	97.6	65.9	63.1	92.0	0.5	66.1	1.8	0.27
0.5, 0.7, 0.9	97.7	99.1	96.5	97.9	66.6	63.9	93.0	1.5	70.9	2.3	0.27
0.7, 0.8, 0.9	96.6	98.7	96.2	97.0	62.6	60.7	91.6	0.6	64.3	1.7	0.28
Filter rate κ											
1	95.2	98.2	95.0	97.2	61.7	59.7	91.4	0.9	66.0	2.0	0.28
2	96.6	98.6	95.9	98.0	65.3	62.4	92.8	1.5	70.8	2.3	0.28
4 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
8	96.8	98.6	96.2	97.8	67.4	64.8	91.6	3.7	74.7	3.9	0.29
16	95.2	97.8	95.2	97.6	65.4	63.3	90.1	3.8	72.7	4.2	0.28
VisA											
Probe paths K_p											
4	92.1	93.3	89.7	98.2	35.4	41.7	87.3	19.0	95.3	23.1	0.16
8	94.8	95.5	92.3	98.6	38.4	44.5	89.3	7.5	92.6	11.3	0.17
16	96.3	96.7	93.8	98.7	39.9	45.8	89.9	2.9	89.5	6.1	0.20
32 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
64	97.1	97.4	94.3	98.7	40.7	46.5	90.0	0.7	84.2	3.3	0.43
128	97.2	97.5	94.5	98.7	40.9	46.6	90.0	0.5	82.3	2.8	0.72
Number of probe times											
0.9	83.4	87.2	82.8	96.0	21.0	28.0	79.5	13.5	90.7	16.2	0.11
0.6, 0.9	96.3	97.1	93.4	98.6	39.0	45.5	90.0	1.7	86.9	4.9	0.19
0.3, 0.6, 0.9 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
0.3, 0.5, 0.7, 0.9	96.7	97.3	93.9	98.2	39.4	45.4	90.5	0.6	81.9	2.9	0.36
0.1, 0.3, . . . , 0.9 (5 times)	96.5	97.0	93.7	98.5	41.8	46.5	90.5	0.7	84.8	3.1	0.45
0.1, 0.2, . . . , 0.9 (9 times)	96.6	97.1	93.6	98.4	41.8	46.5	90.8	0.3	80.4	2.0	0.78
Placement of three probe times											
0.3, 0.6, 0.9 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
0.15, 0.45, 0.9	95.5	96.2	93.0	98.7	41.3	46.8	89.0	3.4	90.5	6.7	0.27
0.2, 0.5, 0.8	96.2	96.7	93.5	98.6	40.3	45.9	89.2	0.3	78.6	1.8	0.28
0.5, 0.7, 0.9	96.8	97.4	94.1	98.4	38.3	45.0	90.5	0.6	81.5	3.0	0.28
0.7, 0.8, 0.9	96.8	97.3	93.7	97.9	36.2	43.1	91.1	0.4	76.7	2.2	0.27
Filter rate κ											
1	94.3	95.3	91.7	98.1	38.4	44.1	87.7	0.4	76.4	1.8	0.28
2	95.4	96.2	92.8	98.5	39.3	45.1	89.1	0.4	81.7	2.4	0.28
4 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
8	96.7	97.3	93.9	98.6	40.4	46.7	90.1	4.4	88.3	7.3	0.28
16	96.2	97.1	93.4	98.5	39.6	46.4	89.8	6.5	88.4	9.1	0.29

Table 11: Ablation of the cycles and the readout (Section 3.4) on MVTec-AD (top) and VisA (bottom), seed 0, in %. The belief columns are defined in Section B.3. The block "Reported map" scores three maps produced by the reference run, with the same post-processing.

	Image-level			Pixel-level				Belief $q > \frac{1}{2}$			
Setting	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO	normal	recall	FRR	s/img
MVTec-AD											
Merge of the cycle evidence											
soft minimum (54) (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
mean	96.4	98.5	95.9	98.1	65.1	62.5	92.7	4.0	81.4	5.1	0.28
maximum	95.3	97.9	95.1	97.9	62.7	60.8	91.8	4.7	85.3	6.8	0.27
last cycle only	95.7	98.2	95.5	98.1	64.5	62.3	92.6	4.4	83.8	6.1	0.27
Probe target of cycles $c > 1$											
previous restoration (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
observation M	97.9	99.1	96.7	98.0	67.0	64.5	92.5	1.4	69.1	1.9	0.28
observation M, sources of cycle 1	97.7	99.1	96.6	98.0	66.9	64.3	92.5	2.1	71.3	2.6	0.28
Sources of cycles $c > 1$											
new draws (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
sources of cycle 1	97.6	99.0	96.5	98.1	67.9	65.1	93.0	3.3	75.4	3.5	0.28
Cycles C											
1	97.6	99.0	96.4	98.0	66.5	64.0	92.8	2.1	71.3	2.6	0.15
2 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
3	97.6	98.9	96.3	98.2	68.2	65.4	93.1	3.3	76.5	3.5	0.42
4	97.5	98.9	96.2	98.2	68.3	65.5	93.0	3.4	77.2	3.7	0.54
Pooling of the K_c *residuals in* (21)											
maximum (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
soft minimum	97.4	98.9	96.5	98.1	67.0	64.7	93.1	3.0	75.1	3.3	0.27
mean energy	97.6	99.0	96.4	98.1	67.3	64.7	93.1	3.0	75.1	3.3	0.28
energy of the mean restoration	97.6	99.0	96.5	98.0	66.8	64.2	93.2	3.0	75.1	3.3	0.28
Reported map											
residual (21) (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	3.0	75.1	3.3	0.28
merged probe evidence $\sqrt{2\bar{S}}$	98.1	99.2	97.1	97.6	68.7	66.0	93.3	3.0	75.1	3.3	0.28
belief q	96.7	98.6	96.1	98.0	70.2	66.3	90.9	3.0	75.1	3.3	0.28
VisA											
Merge of the cycle evidence											
soft minimum (54) (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
mean	95.6	96.1	93.0	98.7	37.9	43.8	89.7	2.7	92.5	6.2	0.28
maximum	94.7	95.3	92.1	98.7	35.5	42.0	89.0	6.3	94.8	10.5	0.28
last cycle only	94.9	95.4	92.7	98.7	38.3	43.8	89.1	4.4	93.7	8.3	0.29
Probe target of cycles $c > 1$											
previous restoration (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
observation M	97.1	97.5	94.4	98.6	40.5	46.3	90.1	0.5	79.8	2.8	0.29
observation M, sources of cycle 1	97.1	97.5	94.4	98.6	40.6	46.4	90.2	1.0	81.6	3.5	0.28
Sources of cycles $c > 1$											
new draws (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
sources of cycle 1	96.8	97.2	94.1	98.7	40.4	46.4	90.0	1.6	86.7	4.5	0.29
Cycles C											
1	96.5	97.0	93.8	98.6	40.0	45.7	89.9	1.0	81.6	3.5	0.15
2 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
3	96.7	97.1	94.1	98.7	40.4	46.2	90.1	1.4	88.7	4.5	0.42
4	96.7	97.0	94.4	98.8	40.3	46.1	90.1	1.5	90.0	4.7	0.54
Pooling of the K_c *residuals in* (21)											
maximum (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
soft minimum	96.9	97.3	94.2	98.8	40.6	46.1	90.6	1.3	86.3	4.2	0.28
mean energy	96.9	97.2	94.1	98.8	40.5	46.3	90.6	1.3	86.3	4.2	0.28
energy of the mean restoration	96.9	97.3	94.3	98.8	40.5	46.2	91.3	1.3	86.3	4.2	0.28
Reported map											
residual (21) (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1	1.3	86.3	4.2	0.29
merged probe evidence $\sqrt{2\bar{S}}$	97.1	97.7	94.3	98.0	39.4	44.8	91.0	1.3	86.3	4.2	0.29
belief q	94.1	94.2	91.4	98.4	37.0	41.2	87.4	1.3	86.3	4.2	0.29

Table 12: Post-processing of the reported map on MVTec-AD and VisA, seed 0, mean over the categories, in %. Each setting is applied to the saved unsmoothed maps of the reference run; the controller is not run again. The image-score rules change only the image-level metrics.

	MVTec-AD							VisA						
	Image-level			Pixel-level				Image-level			Pixel-level			
Setting	AUROC	AP	F_1	AUROC	AP	F_1	PRO	AUROC	AP	F_1	AUROC	AP	F_1	PRO
Kernel size k, $\sigma = 3$														
1 (no smoothing)	97.0	98.8	96.6	96.2	64.5	63.8	89.5	96.8	97.2	94.1	97.8	42.3	49.3	90.4
5	97.5	99.0	96.6	97.9	68.8	66.2	93.1	96.8	97.2	94.1	98.6	42.3	48.6	90.6
7	97.6	99.0	96.5	98.1	68.3	65.7	93.1	96.8	97.2	94.1	98.6	41.2	47.2	90.3
9 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	96.8	97.2	94.3	98.7	40.3	46.2	90.1
13	97.6	99.0	96.4	98.2	67.3	64.3	92.9	96.7	97.1	94.1	98.7	39.6	45.4	89.9
17	97.6	99.0	96.4	98.2	67.2	64.1	92.9	96.7	97.1	94.0	98.8	39.5	45.2	89.9
Standard deviation σ, $k = 9$														
1	97.5	99.0	96.7	97.8	69.2	66.4	93.2	96.8	97.2	94.0	98.5	43.7	49.7	91.3
2	97.6	99.0	96.6	98.1	68.8	65.9	93.3	96.8	97.2	94.1	98.7	41.8	47.8	90.6
3 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	96.8	97.2	94.3	98.7	40.3	46.2	90.1
4.5	97.6	99.0	96.4	98.1	67.1	64.5	92.8	96.7	97.1	94.1	98.7	39.4	45.3	89.7
6	97.6	99.0	96.4	98.1	66.7	64.3	92.7	96.7	97.1	94.0	98.7	39.0	45.0	89.5
Image score														
top-1% mean (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0	96.8	97.2	94.3	98.7	40.3	46.2	90.1
top-0.5% mean	97.4	98.9	96.2	98.1	67.8	65.1	93.0	96.7	97.2	93.7	98.7	40.3	46.2	90.1
top-2% mean	97.5	99.0	96.6	98.1	67.8	65.1	93.0	96.7	97.1	94.2	98.7	40.3	46.2	90.1
top-5% mean	97.0	98.8	96.4	98.1	67.8	65.1	93.0	96.4	96.8	93.8	98.7	40.3	46.2	90.1
maximum	96.6	98.4	95.8	98.1	67.8	65.1	93.0	96.1	96.9	93.3	98.7	40.3	46.2	90.1
$\frac{1}{2}$(maximum + top-1% mean)	97.1	98.7	95.9	98.1	67.8	65.1	93.0	96.4	97.0	93.5	98.7	40.3	46.2	90.1

Table 13: The reference configuration with 16 seeds on MVTec-AD and VisA, in %, mean over the categories; s/img in seconds. The last two rows of each block are the mean and the sample standard deviation over the 16 runs. The belief columns are defined in Section B.3.

	Image-level			Pixel-level				Belief $q > \frac{1}{2}$			
Seed	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO	normal	recall	FRR	s/img
MVTec-AD											
0	97.66	99.02	96.52	98.12	67.85	65.07	93.04	3.03	75.06	3.27	0.285
1	97.78	99.07	96.95	98.14	67.87	65.11	93.01	3.07	74.99	3.28	0.276
2	97.71	99.02	96.71	98.13	67.92	65.26	93.10	3.05	75.07	3.31	0.275
3	97.56	98.98	96.45	98.11	67.92	65.22	93.06	3.10	75.12	3.29	0.275
4	97.76	99.07	96.76	98.14	68.05	65.24	93.01	3.04	75.08	3.25	0.279
5	97.53	98.98	96.45	98.13	67.90	65.15	93.05	3.06	75.03	3.29	0.276
6	97.67	99.03	96.36	98.12	67.94	65.22	92.95	3.04	75.03	3.28	0.282
7	97.66	99.02	96.59	98.13	67.93	65.23	92.93	3.02	75.20	3.25	0.279
8	97.65	99.03	96.59	98.13	67.91	65.11	93.01	3.04	75.21	3.28	0.281
9	97.72	99.03	96.50	98.12	67.88	65.18	92.97	3.03	74.91	3.27	0.277
10	97.57	98.98	96.45	98.11	67.88	65.18	92.98	3.09	75.06	3.30	0.281
11	97.63	98.98	96.62	98.13	67.84	65.11	92.94	3.06	75.15	3.27	0.282
12	97.78	99.05	96.85	98.13	67.90	65.16	92.96	2.96	75.09	3.26	0.280
13	97.82	99.07	96.91	98.13	67.84	65.16	92.96	3.14	74.91	3.29	0.285
14	97.69	99.02	96.69	98.13	67.85	65.17	92.97	3.06	75.14	3.26	0.289
15	97.75	99.02	96.60	98.13	67.85	65.15	93.05	3.05	75.07	3.28	0.278
mean	97.68	99.02	96.63	98.13	67.89	65.17	93.00	3.05	75.07	3.28	0.280
sd	0.08	0.03	0.17	0.01	0.05	0.05	0.05	0.04	0.09	0.02	0.004
VisA											
0	96.76	97.16	94.31	98.69	40.32	46.19	90.06	1.29	86.32	4.16	0.291
1	97.00	97.27	94.44	98.70	40.22	46.24	90.11	1.28	86.67	4.15	0.282
2	96.87	97.16	94.35	98.69	40.49	46.26	90.06	1.27	86.48	4.15	0.286
3	96.94	97.28	94.34	98.70	40.31	46.19	90.08	1.28	86.42	4.15	0.275
4	96.79	97.16	94.13	98.68	40.45	46.16	90.03	1.28	86.66	4.16	0.283
5	96.86	97.23	94.17	98.69	40.31	46.39	90.07	1.27	86.34	4.15	0.278
6	96.83	97.19	94.19	98.70	40.25	46.08	90.10	1.28	86.42	4.15	0.275
7	96.86	97.12	94.39	98.69	40.22	46.12	90.07	1.27	86.29	4.15	0.282
8	96.89	97.20	94.29	98.68	40.20	46.18	90.01	1.30	86.46	4.16	0.279
9	96.96	97.18	94.56	98.71	40.42	46.22	90.11	1.27	86.85	4.15	0.283
10	96.87	97.22	94.40	98.70	40.19	46.12	90.09	1.28	86.53	4.16	0.285
11	96.76	97.09	94.36	98.69	40.22	46.16	90.17	1.29	86.25	4.17	0.283
12	96.95	97.38	94.43	98.70	40.46	46.19	90.07	1.29	86.75	4.15	0.279
13	96.96	97.30	94.36	98.70	40.55	46.31	90.10	1.27	86.40	4.14	0.287
14	96.88	97.15	94.18	98.69	40.12	46.04	90.10	1.30	86.46	4.15	0.291
15	96.92	97.22	94.38	98.71	40.20	46.12	90.13	1.28	86.22	4.14	0.280
mean	96.88	97.21	94.33	98.70	40.31	46.19	90.09	1.28	86.47	4.15	0.282
sd	0.07	0.07	0.11	0.01	0.13	0.09	0.04	0.01	0.18	0.01	0.005

B.4 SEEDS

The reference configuration is run with 16 seeds, 0 to 15, on both datasets; a seed changes only the Gaussian source draws of the probe and controlled paths. Table 13 lists every run, and Tables 14 and 15 give the mean and the standard deviation over the 16 runs for every category. The standard deviation of a single category's metric reaches 0.83 points on MVTec-AD (image F_1-max of transistor) and 0.84 points on VisA (image F_1-max of macaroni2), against at most 0.18 and 0.13 points for the means over the categories. Tables 16 and 17 give the mean and the standard deviation over seeds 0, 1 and 2 of the rows of Table 2, of the threshold, temperature and coupling runs of the belief, and of the points of Figure 3.

Table 14: Per-category results of the reference configuration on MVTec-AD (top) and VisA (bottom), mean ± standard deviation over 16 seeds, in %. The row "mean" of each half is the mean over the categories, as in Table 1.

	Image-level			Pixel-level			
Category	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO
MVTec-AD							
bottle	98.31±0.26	99.53±0.07	97.57±0.02	98.14±0.01	78.41±0.06	71.55±0.12	94.20±0.05
cable	92.45±0.26	95.96±0.12	87.98±0.52	97.00±0.03	57.23±0.23	56.35±0.20	86.47±0.26
capsule	95.46±0.23	99.06±0.04	96.18±0.20	97.79±0.07	43.42±0.14	47.47±0.13	94.32±0.08
carpet	92.59±0.47	97.76±0.14	91.90±0.27	97.33±0.05	71.63±0.28	67.98±0.25	85.88±0.55
grid	100.00±0.00	100.00±0.00	100.00±0.00	99.35±0.01	54.30±0.26	54.79±0.20	97.07±0.04
hazelnut	99.68±0.22	99.81±0.14	98.59±0.25	99.39±0.01	78.91±0.13	72.78±0.17	91.33±0.13
leather	99.88±0.08	99.96±0.03	99.49±0.14	99.70±0.00	62.51±0.12	63.98±0.12	98.69±0.01
metal_nut	96.82±0.39	99.19±0.11	97.23±0.41	98.14±0.02	87.45±0.08	81.09±0.10	94.02±0.21
pill	98.45±0.46	99.72±0.09	98.04±0.33	99.36±0.00	82.00±0.11	77.07±0.14	97.99±0.02
screw	95.80±0.44	97.99±0.28	96.22±0.47	98.96±0.04	45.68±0.44	50.55±0.43	95.41±0.14
tile	99.17±0.16	99.68±0.06	97.54±0.59	97.92±0.02	79.92±0.06	72.25±0.17	88.58±0.53
toothbrush	99.83±0.14	99.93±0.05	98.98±0.82	99.30±0.01	72.87±0.33	67.95±0.29	95.13±0.11
transistor	97.81±0.12	97.05±0.17	92.80±0.83	94.56±0.05	60.74±0.17	56.02±0.22	88.78±0.10
wood	99.18±0.19	99.74±0.06	97.91±0.51	95.78±0.05	70.51±0.19	65.51±0.14	90.22±0.14
zipper	99.82±0.05	99.95±0.01	98.98±0.26	99.16±0.01	72.84±0.18	72.22±0.12	96.91±0.06
mean	97.68±0.08	99.02±0.03	96.63±0.17	98.13±0.01	67.89±0.05	65.17±0.05	93.00±0.05
VisA							
candle	95.64±0.23	96.14±0.19	88.95±0.39	99.13±0.04	35.32±0.37	40.24±0.24	97.32±0.13
capsules	92.37±0.33	95.33±0.19	88.77±0.43	99.87±0.00	71.22±0.56	74.62±0.32	97.04±0.13
cashew	95.59±0.55	97.59±0.31	94.24±0.66	97.68±0.06	69.84±0.23	66.11±0.34	73.04±0.33
chewinggum	96.84±0.41	98.53±0.19	94.16±0.45	97.05±0.06	16.72±0.09	35.09±0.11	59.30±0.25
fryum	97.26±0.20	98.51±0.11	96.03±0.46	98.11±0.01	49.84±0.19	54.08±0.16	93.55±0.09
macaroni1	98.42±0.05	97.37±0.09	97.71±0.43	99.86±0.00	22.61±0.54	34.84±0.61	98.71±0.04
macaroni2	93.41±0.30	91.69±0.57	88.74±0.84	99.84±0.01	19.66±0.48	27.08±0.73	98.85±0.04
pcb1	98.95±0.14	98.83±0.20	96.21±0.38	98.68±0.03	43.55±0.55	45.15±0.34	94.52±0.18
pcb2	99.80±0.04	99.81±0.04	98.07±0.36	97.27±0.05	15.44±0.51	26.07±0.71	90.11±0.09
pcb3	97.93±0.09	96.71±0.14	97.03±0.12	99.16±0.03	17.99±0.20	31.68±0.08	95.38±0.04
pcb4	98.67±0.04	97.01±0.23	97.18±0.21	98.11±0.02	45.98±0.46	50.08±0.22	88.62±0.12
pipe_fryum	97.67±0.20	98.96±0.09	94.85±0.34	99.57±0.00	75.54±0.11	69.19±0.15	94.60±0.13
mean	96.88±0.07	97.21±0.07	94.33±0.11	98.70±0.01	40.31±0.13	46.19±0.09	90.09±0.04

Table 15: The final belief at $q > \frac{1}{2}$ per category, mean ± standard deviation over 16 seeds, in %. The columns are defined in Section B.3.

MVTec-AD				VisA			
Category	normal	recall	FRR	Category	normal	recall	FRR
bottle	0.59±0.03	78.48±0.21	1.57±0.01	candle	0.08±0.00	62.95±1.35	0.23±0.00
cable	6.02±0.05	84.93±0.23	8.18±0.05	capsules	2.07±0.04	99.63±0.08	4.18±0.04
capsule	0.08±0.00	63.30±0.24	0.50±0.01	cashew	0.33±0.00	83.56±0.97	2.25±0.03
carpet	36.12±0.57	84.06±0.23	18.33±0.18	chewinggum	0.07±0.00	90.16±0.18	1.65±0.01
grid	0.44±0.02	97.84±0.13	9.53±0.07	fryum	0.23±0.01	81.74±0.41	2.97±0.04
hazelnut	0.51±0.02	93.91±0.15	0.93±0.02	macaroni1	0.08±0.00	90.12±0.80	0.60±0.00
leather	0.21±0.01	97.25±0.15	1.08±0.01	macaroni2	0.16±0.00	84.35±0.97	0.24±0.00
metal_nut	0.72±0.04	86.20±0.20	1.64±0.02	pcb1	2.81±0.06	90.96±0.68	7.06±0.05
pill	0.05±0.00	71.13±0.32	0.43±0.01	pcb2	3.68±0.07	94.35±0.11	12.28±0.04
screw	0.06±0.00	44.14±0.54	0.12±0.00	pcb3	3.82±0.05	96.01±0.16	7.02±0.05
tile	0.07±0.01	64.56±0.19	0.78±0.01	pcb4	2.01±0.03	93.35±0.31	11.07±0.04
toothbrush	0.44±0.04	95.01±0.13	2.63±0.06	pipe_fryum	0.04±0.00	70.47±0.63	0.29±0.01
transistor	0.36±0.01	41.95±0.33	1.15±0.02				
wood	0.08±0.02	69.91±0.30	2.05±0.03				
zipper	0.04±0.00	53.37±0.32	0.26±0.00				
mean	3.05±0.04	75.07±0.09	3.28±0.02	mean	1.28±0.01	86.47±0.18	4.15±0.01

Table 16: Mean ± standard deviation over seeds 0, 1 and 2 of the rows of Table 2, on MVTec-AD (top) and VisA (bottom), in %.

	Image-level			Pixel-level			
Setting	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO
MVTec-AD							
reference	97.71±0.06	99.04±0.03	96.73±0.21	98.13±0.01	67.88±0.04	65.15±0.10	93.05±0.04
no coupling, $\mu = 0$	94.95±0.16	98.00±0.06	94.77±0.17	96.72±0.01	61.31±0.05	59.20±0.05	91.10±0.06
blend at the last step	96.83±0.11	98.85±0.04	96.25±0.15	97.29±0.01	62.90±0.04	60.02±0.06	92.01±0.01
mean energy over sources	76.61±0.18	88.49±0.09	87.92±0.09	90.35±0.03	33.20±0.18	36.64±0.11	74.85±0.14
last measurement only	95.13±0.12	97.80±0.05	95.06±0.06	97.55±0.02	65.30±0.09	63.29±0.14	89.82±0.02
mean merge over cycles	96.51±0.10	98.49±0.03	95.94±0.08	98.08±0.00	65.10±0.22	62.57±0.23	92.66±0.04
$C = 1$, with control	97.66±0.05	99.00±0.02	96.48±0.11	98.03±0.01	66.48±0.05	63.91±0.06	92.80±0.04
$C = 1$, no control	64.11±0.37	83.00±0.26	85.31±0.10	80.39±0.10	20.44±0.04	25.35±0.12	50.26±0.22
$N = 10$, probe paths	97.78±0.02	99.12±0.02	96.93±0.12	97.58±0.00	65.86±0.02	63.76±0.04	92.78±0.04
$N = 10$, controlled path	71.64±0.54	85.58±0.23	87.82±0.15	88.69±0.08	27.31±0.09	31.28±0.05	67.62±0.19
VisA							
reference	96.88±0.12	97.20±0.07	94.37±0.07	98.69±0.00	40.35±0.14	46.23±0.03	90.08±0.03
no coupling, $\mu = 0$	92.74±0.05	94.14±0.05	89.89±0.08	97.62±0.02	37.94±0.22	44.48±0.14	87.85±0.05
blend at the last step	96.90±0.06	97.41±0.04	93.90±0.15	98.15±0.01	37.17±0.10	43.58±0.11	91.69±0.07
mean energy over sources	71.96±0.31	76.75±0.19	77.35±0.14	92.76±0.02	15.54±0.03	19.66±0.04	71.42±0.07
last measurement only	96.28±0.11	97.16±0.08	93.43±0.05	98.46±0.01	39.36±0.09	46.19±0.07	89.78±0.05
mean merge over cycles	95.55±0.05	96.05±0.04	93.03±0.10	98.73±0.01	37.89±0.26	43.90±0.11	89.76±0.04
$C = 1$, with control	96.46±0.01	96.98±0.03	93.82±0.09	98.58±0.01	39.95±0.22	45.44±0.31	89.92±0.03
$C = 1$, no control	62.63±0.43	68.30±0.31	74.75±0.08	86.26±0.13	7.47±0.07	10.72±0.12	55.87±0.27
$N = 10$, probe paths	96.99±0.15	97.42±0.04	93.93±0.37	97.91±0.01	39.58±0.11	45.42±0.04	90.92±0.02
$N = 10$, controlled path	65.02±0.32	70.16±0.36	75.35±0.26	89.69±0.08	8.93±0.15	11.99±0.16	61.51±0.05

Table 17: Mean ± standard deviation over seeds 0, 1 and 2, in %, of the threshold, temperature and coupling runs of the belief and of the points of Figure 3, on MVTec-AD (top) and VisA (bottom).

	Image-level			Pixel-level			
Setting	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO
MVTec-AD							
Belief, Table 5							
$\lambda = 0.045$ ($T = 0.015$)	95.21±0.19	97.80±0.08	94.91±0.09	97.89±0.02	65.07±0.12	62.62±0.08	92.04±0.03
$\lambda = 0.125$ ($T = 0.025$)	97.34±0.06	99.00±0.01	96.61±0.09	97.95±0.01	66.11±0.04	63.85±0.03	92.11±0.06
$\lambda = 0.180$	95.35±0.12	98.39±0.04	95.52±0.11	97.61±0.01	60.87±0.06	59.05±0.05	92.12±0.07
$T = 0.010$	97.55±0.13	99.06±0.04	96.38±0.15	98.03±0.01	66.77±0.03	63.82±0.06	93.22±0.06
$T = 0.030$	97.26±0.05	98.82±0.03	96.35±0.20	97.92±0.01	67.07±0.05	64.42±0.08	92.56±0.06
$T = 0.040$	96.58±0.10	98.57±0.03	95.77±0.15	97.59±0.01	65.89±0.06	63.42±0.06	91.88±0.06
$\mu/\lambda = 0.25$	96.26±0.16	98.47±0.05	95.62±0.18	97.52±0.01	64.22±0.05	61.48±0.10	92.66±0.05
$\mu/\lambda = 0.5$	97.09±0.09	98.77±0.03	96.22±0.31	97.89±0.01	66.09±0.06	63.21±0.10	93.18±0.01
$\mu/\lambda = 2$	97.27±0.06	98.99±0.03	96.60±0.06	98.04±0.01	67.94±0.01	65.96±0.05	91.16±0.08
Budget, Figure 3							
$K_p = 4$	90.72±0.43	95.01±0.20	93.14±0.28	97.09±0.01	59.68±0.09	58.65±0.04	90.04±0.04
$K_p = 8$	95.44±0.10	97.88±0.03	95.31±0.16	97.87±0.01	65.43±0.03	62.92±0.05	92.30±0.04
$K_p = 16$	97.22±0.07	98.81±0.03	96.23±0.15	98.07±0.01	67.32±0.01	64.52±0.06	92.90±0.07
$K_p = 64$	97.82±0.04	99.09±0.02	96.71±0.17	98.13±0.01	67.97±0.05	65.34±0.07	92.92±0.12
$K_p = 128$	97.87±0.09	99.13±0.03	96.71±0.08	98.13±0.00	67.90±0.05	65.36±0.00	92.71±0.05
$J = 1$	85.37±0.25	92.69±0.06	91.50±0.09	95.32±0.02	51.37±0.16	52.05±0.19	81.73±0.12
$J = 2$	97.09±0.13	98.69±0.04	96.42±0.30	97.95±0.01	66.02±0.12	63.57±0.17	91.93±0.06
$J = 4$	97.88±0.05	99.14±0.02	96.71±0.09	97.87±0.00	67.12±0.04	64.33±0.04	93.20±0.03
$J = 5$	97.52±0.09	98.97±0.03	96.32±0.24	97.97±0.01	67.36±0.03	64.27±0.03	93.13±0.05
$J = 9$	97.49±0.11	99.02±0.04	96.49±0.15	97.76±0.01	65.81±0.01	62.69±0.02	92.96±0.03
$C = 3$	97.62±0.06	98.97±0.03	96.56±0.24	98.17±0.01	68.16±0.03	65.40±0.05	93.15±0.09
$C = 4$	97.56±0.04	98.94±0.02	96.51±0.26	98.19±0.01	68.33±0.01	65.53±0.01	93.11±0.09
VisA							
Belief, Table 5							
$\lambda = 0.045$ ($T = 0.015$)	93.45±0.06	94.45±0.04	90.97±0.04	98.31±0.02	36.40±0.19	42.60±0.17	87.64±0.06
$\lambda = 0.125$ ($T = 0.025$)	97.22±0.12	97.66±0.10	94.25±0.17	98.69±0.00	40.83±0.09	46.42±0.11	90.90±0.06
$\lambda = 0.180$	96.24±0.22	97.23±0.13	92.87±0.28	98.35±0.01	36.90±0.06	43.65±0.04	91.86±0.05
$T = 0.010$	96.85±0.10	97.29±0.07	93.96±0.13	98.56±0.01	38.70±0.09	45.07±0.07	91.55±0.02
$T = 0.030$	96.09±0.03	96.76±0.04	93.44±0.16	98.50±0.01	40.63±0.17	46.59±0.08	89.32±0.03
$T = 0.040$	95.14±0.06	96.27±0.07	92.72±0.09	98.30±0.02	40.58±0.19	46.53±0.10	88.91±0.05
$\mu/\lambda = 0.25$	94.70±0.04	95.52±0.01	92.14±0.03	98.17±0.02	38.59±0.18	44.95±0.07	89.49±0.02
$\mu/\lambda = 0.5$	95.81±0.05	96.36±0.02	93.34±0.15	98.47±0.01	39.21±0.14	45.43±0.04	90.07±0.06
$\mu/\lambda = 2$	97.31±0.08	97.62±0.09	94.49±0.09	98.72±0.00	41.68±0.08	47.51±0.09	89.31±0.02
Budget, Figure 3							
$K_p = 4$	91.90±0.20	93.05±0.18	89.31±0.39	98.21±0.02	35.31±0.18	41.51±0.20	87.33±0.01
$K_p = 8$	94.96±0.16	95.60±0.10	92.34±0.03	98.58±0.01	38.33±0.10	44.47±0.05	89.27±0.02
$K_p = 16$	96.32±0.05	96.72±0.07	93.81±0.18	98.68±0.01	39.77±0.13	45.66±0.12	89.86±0.01
$K_p = 64$	97.18±0.07	97.43±0.12	94.49±0.14	98.70±0.00	40.71±0.10	46.54±0.03	90.07±0.03
$K_p = 128$	97.27±0.06	97.57±0.03	94.58±0.12	98.69±0.02	40.73±0.18	46.47±0.12	89.97±0.09
$J = 1$	83.43±0.28	87.16±0.10	82.86±0.23	96.02±0.03	21.02±0.05	27.96±0.17	79.49±0.03
$J = 2$	96.24±0.07	97.03±0.07	93.36±0.06	98.60±0.01	39.01±0.04	45.57±0.10	90.01±0.02
$J = 4$	96.81±0.06	97.25±0.04	93.97±0.06	98.17±0.02	39.57±0.12	45.41±0.06	90.49±0.01
$J = 5$	96.55±0.07	96.99±0.08	93.77±0.06	98.53±0.02	41.85±0.07	46.62±0.09	90.46±0.00
$J = 9$	96.66±0.06	97.13±0.08	93.66±0.06	98.40±0.02	41.77±0.03	46.52±0.01	90.83±0.03
$C = 3$	96.77±0.02	97.03±0.09	94.27±0.16	98.74±0.00	40.44±0.06	46.24±0.03	90.04±0.04
$C = 4$	96.66±0.09	96.93±0.11	94.34±0.04	98.78±0.00	40.19±0.13	45.98±0.14	90.01±0.07

Table 18: The reference configuration on the checkpoints saved during training, seed 0, mean over the categories, in %. The best value of each column is in bold; all other results use the checkpoint at 2,000 epochs.

	MVTec-AD							VisA						
	Image-level			Pixel-level				Image-level			Pixel-level			
Epochs	AUROC	AP	F_1	AUROC	AP	F_1	PRO	AUROC	AP	F_1	AUROC	AP	F_1	PRO
100	89.8	95.6	93.1	96.4	58.7	57.1	88.0	94.7	95.4	91.8	98.2	38.2	44.7	88.5
200	93.3	97.2	94.1	96.9	61.7	60.1	90.3	95.8	96.3	93.1	98.2	38.0	44.5	89.0
300	94.6	97.6	94.8	97.4	63.9	61.9	91.3	96.2	96.7	93.4	98.5	39.5	45.7	89.6
400	94.9	97.8	94.9	97.4	64.2	62.3	91.4	96.3	96.8	93.4	98.4	39.5	45.8	89.3
500	95.9	98.3	95.8	97.6	65.3	63.5	92.0	96.5	96.9	93.8	98.5	39.7	46.0	89.8
600	96.0	98.3	95.7	97.6	65.6	63.4	91.9	96.5	96.9	93.8	98.5	39.3	45.7	89.6
700	96.0	98.3	95.6	97.7	65.9	63.7	92.2	96.5	96.9	93.8	98.6	40.3	46.3	89.9
800	96.8	98.7	96.0	97.9	66.7	64.3	92.5	96.7	97.0	94.0	98.6	39.8	46.1	89.9
900	97.0	98.8	96.1	97.8	66.2	64.1	92.3	96.6	97.0	93.9	98.6	40.3	46.5	90.0
1000	96.9	98.7	96.0	97.9	66.6	64.2	92.5	96.8	97.2	94.1	98.6	40.5	46.4	90.0
1100	97.2	98.8	96.4	97.9	66.7	64.4	92.5	**96.9**	**97.3**	94.0	98.6	40.2	46.2	90.0
1200	97.3	98.9	96.3	97.9	67.1	64.8	92.7	96.7	97.2	93.9	**98.7**	40.8	**46.8**	90.1
1300	97.3	98.9	96.3	98.0	67.5	65.0	92.8	**96.9**	**97.3**	94.1	98.6	40.5	46.4	90.1
1400	97.2	98.9	96.2	98.0	67.4	64.9	92.7	**96.9**	97.2	94.2	**98.7**	40.6	46.3	**90.2**
1500	97.4	99.0	96.3	98.0	67.6	65.2	92.9	**96.9**	97.2	**94.3**	98.6	40.1	45.9	89.8
1600	**97.8**	**99.1**	**96.6**	98.0	**68.0**	**65.5**	92.9	96.8	97.2	94.0	**98.7**	**41.2**	46.7	90.1
1700	97.6	99.0	96.5	98.0	67.4	64.9	92.8	96.8	97.2	94.1	98.6	39.9	46.0	89.9
1800	97.6	99.0	96.4	97.9	67.3	64.6	92.7	96.8	97.2	94.1	98.6	40.6	46.4	90.0
1900	97.6	99.0	96.5	**98.1**	67.8	65.0	92.8	**96.9**	**97.3**	94.1	**98.7**	40.7	46.5	89.9
2000 (default)	97.7	99.0	96.5	**98.1**	67.8	65.1	**93.0**	96.8	97.2	**94.3**	**98.7**	40.3	46.2	90.1

B.5 Training epochs

The flow model of each dataset is saved every 100 epochs of training. We run the reference configuration at seed 0 on every checkpoint, from 100 to 2,000 epochs: Table 18 lists the means over the categories, Figure 5 plots image AUROC, pixel AP and AUPRO of these means, and Figure 6 plots image AUROC, pixel AUROC and AUPRO for every category. All other results use the checkpoint at 2,000 epochs. On MVTec-AD, image AUROC rises from 89.8 at 100 epochs to 96.9 at 1,000, and pixel AP from 58.7 to 66.6; from 1,200 epochs on, image AUROC stays within 97.2 to 97.8 and pixel AP within 67.1 to 68.0. On VisA, from 1,000 epochs on, image AUROC stays within 96.7 to 96.9 and pixel AP within 39.9 to 41.2. The 2,000-epoch checkpoint is within 0.1 points of the best checkpoint in image AUROC on both datasets, and within 0.2 (MVTec-AD) and 0.9 (VisA) points in pixel AP. The categories whose image AUROC rises most between 100 and 2,000 epochs are screw (47.0 to 96.5), metal nut (77.2 to 96.6), capsule (78.1 to 95.6) and cable (81.5 to 92.1) on MVTec-AD, and macaroni2 (87.0 to 93.5) and candle (91.0 to 95.4) on VisA.

B.6 Baselines

All baselines of Table 1 except Flow Mismatching operate on the features of a pretrained network: ImageNet-pretrained CNNs (WideResNet-50 for RD4AD, CFLOW-AD, RD++, SimpleNet, ReContrast and InvAD; ResNet-18 for DeSTSeg; ResNet-34 for MambaAD; EfficientNet for UniAD, HVQ-Trans, AMDM and OmiAD), self-supervised ViTs (DINO ViT-S for ViTAD, DINOv2-R ViT-B for Dinomaly and INP-Former), or a pretrained Stable Diffusion autoencoder and denoiser with a ResNet-50 scorer (DiAD). SimpleNet, DeSTSeg and RD++ additionally train on synthetic anomalies. Most baselines take 256×256 inputs, and Dinomaly and INP-Former 392×392. FlowCTRL, like Flow Mismatching, works in pixel space at 128×128; its only learned component is the flow model trained on the normal training images, and it uses no pretrained feature extractor, synthetic anomalies or feature memory.

B.7 Computational cost

Table 19 lists the parameters, the FLOPs and the throughput of the methods of Table 1 except ReContrast. They differ in input resolution, in whether a frozen backbone is counted, in the hardware and in the batch size, so they are not strictly comparable. FlowCTRL uses the same network of Flow Mismatching, a U-Net with 45.3M parameters, and its cost per image is set by the number of network

evaluations (22): the reference configuration evaluates 224 states of 128×128 pixels per image in 8 batched calls, and Flow Mismatching evaluates 100 at its default $K = T = 10$. For these two methods, Table 19 gives the FLOPs of one evaluation and of one image, counted with the FLOP counter of PyTorch: for FlowCTRL during one run of each configuration, for Flow Mismatching as 100 evaluations of the same network. FLOPs count the multiply-accumulates of the network, as in the other rows; the belief solver and the edits of the controller are elementwise and are not counted. The FPS of FlowCTRL is the inverse of its controller time on one NVIDIA A100 GPU with `torch.compile` and bf16 (Section 4.1), which change each metric by less than 0.15 points (Table 4); without compilation, the controller takes 0.68 s per image. With one cycle, FlowCTRL evaluates 112 states per image at 6.88 images per second; Table 2 gives its accuracy.

Table 19: Parameters, FLOPs and throughput of the methods of Table 1. FLOPs count multiply-accumulates; for the two flow methods they are given per network evaluation and per image. Baseline values are reported ones and are not strictly comparable (Section B.7).

Method	Params	FLOPs	FPS
RD4AD (Deng & Li, 2022)	80.6M	28.4G	4.82
CFLOW-AD (Gudovskiy et al., 2022)	237M	28.7G	30.8
UniAD (You et al., 2022)	24.5M	3.4G–3.6G	5.29
RD++ (Tien et al., 2023)	–	–	–[a]
SimpleNet (Liu et al., 2023c)	72.8M	16.1G–17.7G	4.93
DeSTSeg (Zhang et al., 2023)	35.2M	30.7G–122.7G	90.8
HVQ-Trans (Lu et al., 2023)	18.0M[b]	7.38G[b]	5.57
DiAD (He et al., 2024b)	1331M	451.5G–>2.2T	1.64
MambaAD (He et al., 2024a)	25.7M	8.3G	50.7
ViTAD (Zhang et al., 2025)	39.0M	9.7G	115.0
InvAD (Zhang et al., 2026)	95.6M	45.4G	83.5[c]
AMDM / OmiAD (Feng et al., 2025)	–	–	39.37
Dinomaly (Guo et al., 2025)	148.0M	104.7G	58.1[d]
INP-Former (Luo et al., 2025)	139.8M	98.0G	–
Flow Mismatching (Chen et al., 2026)	45.3M	25.3G / 2.53T	3.87
FlowCTRL, one cycle ($C = 1$)	45.3M	25.3G / 2.83T	6.88
FlowCTRL (reference)	45.3M	25.3G / 5.67T	3.57

[a] Only relative speed is reported: 6× PatchCore and 2× CFA. [b] Without the backbone. [c] V100, batch size 32, 256×256 pixels. [d] RTX 3090, batch size 16, 392×392 pixels.

B.8 Evaluation at other resolutions

FlowCTRL runs at 128×128 pixels, while most baselines of Table 1 are evaluated at 256×256, and Dinomaly and INP-Former at 392×392 (Section B.6). Training a flow model in pixel space beyond 128×128 is no longer a trivial task, so we evaluate the maps of the 128×128 model at the larger sizes instead. The final map of every test image, after the smoothing of Section 3.4, is resized bilinearly to $S \times S$ pixels, $S \in \{224, 256, 392\}$, and not smoothed again; the ground-truth mask is resized from its original file to the same size, by nearest neighbour as for $S = 128$. The image score is recomputed from the resized map by the same rule, and the seven metrics are those of Table 1, on the reference run with seed 0; at $S = 128$ this reproduces the run's own evaluation exactly. Table 20 gives the results. At every size, the image-level metrics are unchanged, pixel AUROC and AP change by at most 0.1 points, and pixel F_1-max decreases by at most 0.2 points. AUPRO, which averages the overlap over the connected defect regions of the masks, rises from 93.0 to 93.9–94.1 on MVTec-AD and falls from 90.1 to 88.5–88.6 on VisA.

Table 20: The reference run (seed 0) evaluated at other resolutions: the final map of the 128×128 model resized bilinearly to $S \times S$ pixels, and the ground truth resized from the original masks. Mean over the categories, in %. $S = 128$ is the resolution of the run and of all other tables.

	Image-level			Pixel-level			
Size S	AUROC	AP	F_1-max	AUROC	AP	F_1-max	AUPRO
MVTec-AD							
128 (default)	97.7	99.0	96.5	98.1	67.8	65.1	93.0
224	97.7	99.0	96.5	98.1	67.8	65.0	93.9
256	97.7	99.0	96.5	98.1	67.8	65.0	94.0
392	97.7	99.0	96.5	98.1	67.8	65.0	94.1
VisA							
128 (default)	96.8	97.2	94.3	98.7	40.3	46.2	90.1
224	96.8	97.2	94.3	98.7	40.4	46.1	88.6
256	96.8	97.2	94.3	98.7	40.3	46.0	88.6
392	96.8	97.2	94.3	98.7	40.3	46.0	88.5

APPENDIX C VISUALIZATION

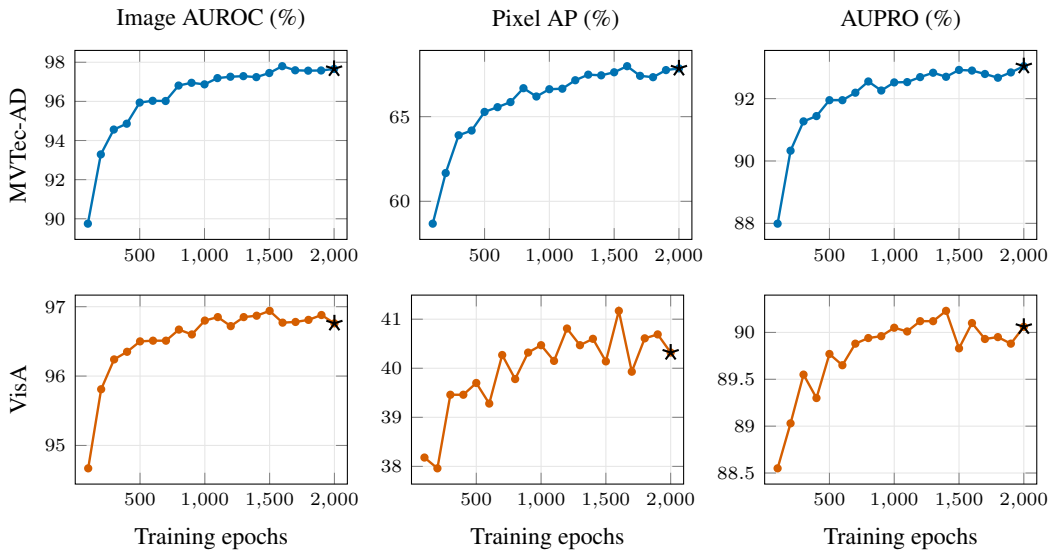


Figure 5: The reference configuration on the checkpoints saved every 100 epochs of training, seed 0, mean over the categories (Table 18). Top: MVTec-AD; bottom: VisA. The star is the checkpoint at 2,000 epochs that all other results use.

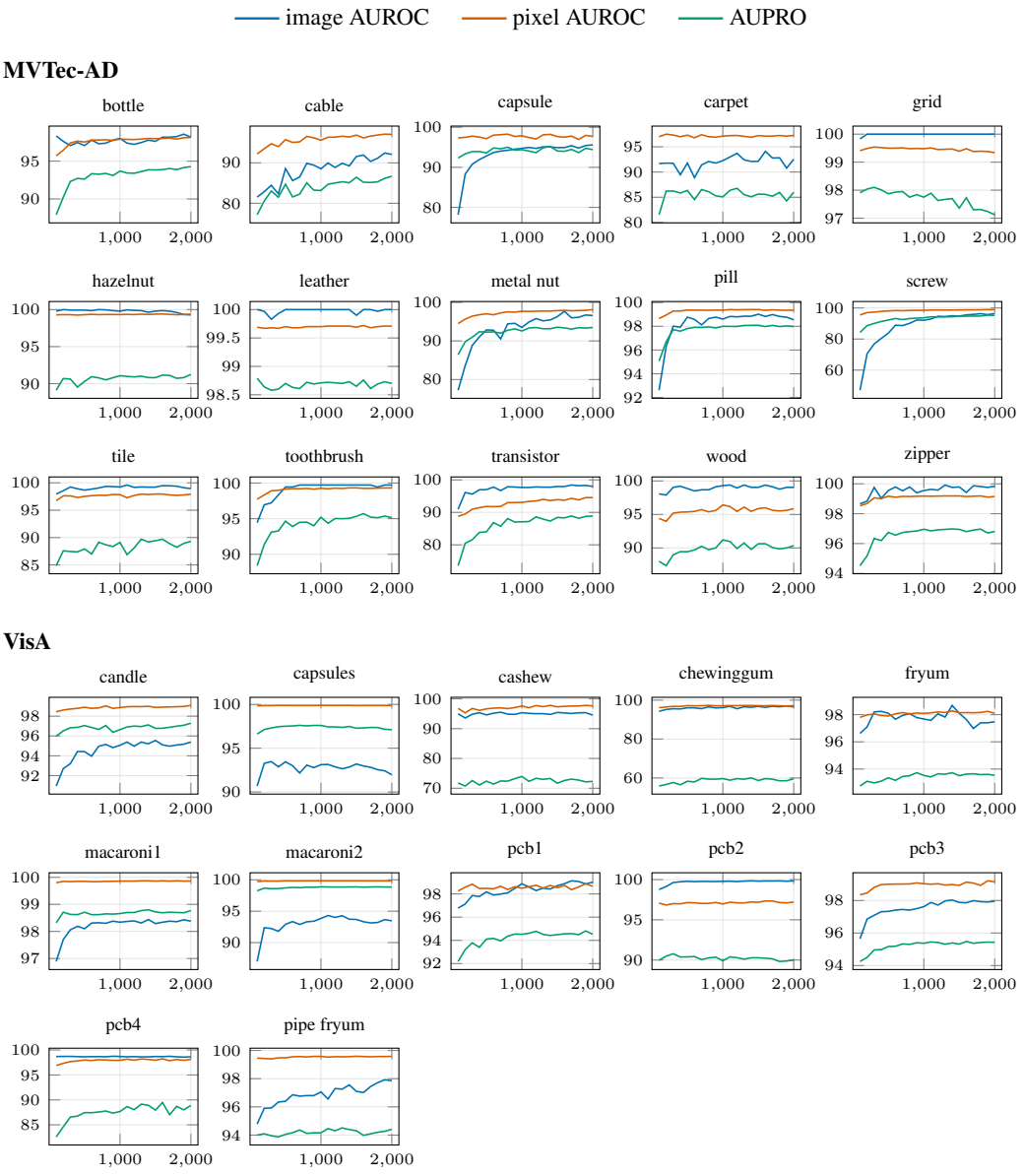


Figure 6: Per-category image AUROC, pixel AUROC and AUPRO (in %) of the reference configuration on the checkpoints saved every 100 epochs of training, seed 0, against the number of epochs, for the 15 categories of MVTec-AD and the 12 of VisA. Table 18 and Figure 5 give the means over the categories.

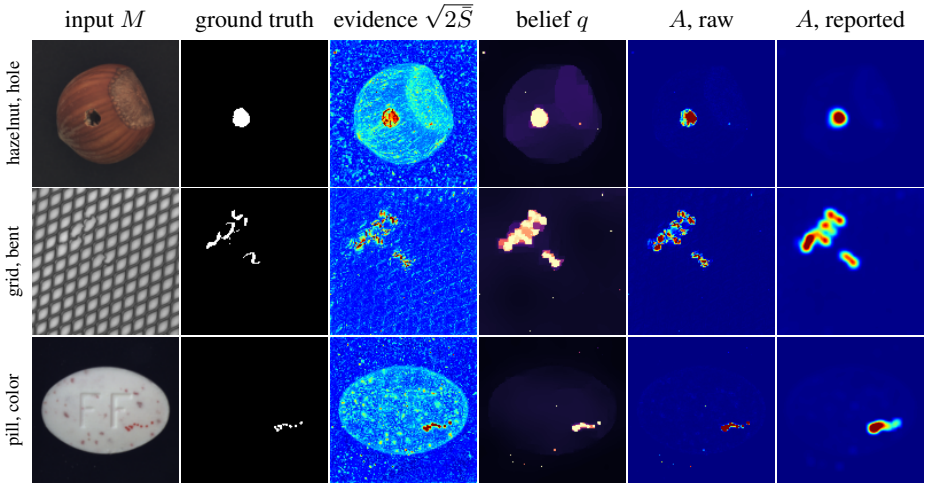


Figure 7: Candidate anomaly maps for the test images of Figure 2: the merged probe evidence $\sqrt{2\bar{S}}$ (54) of the observer, the final belief q, and the residual A of the restorations (21) before and after the 9×9 Gaussian smoothing; the smoothed A is the reported map. The block "Reported map" of Table 11 scores the evidence, the belief and the residual, each after the same smoothing. Colour scales as in Figure 8.

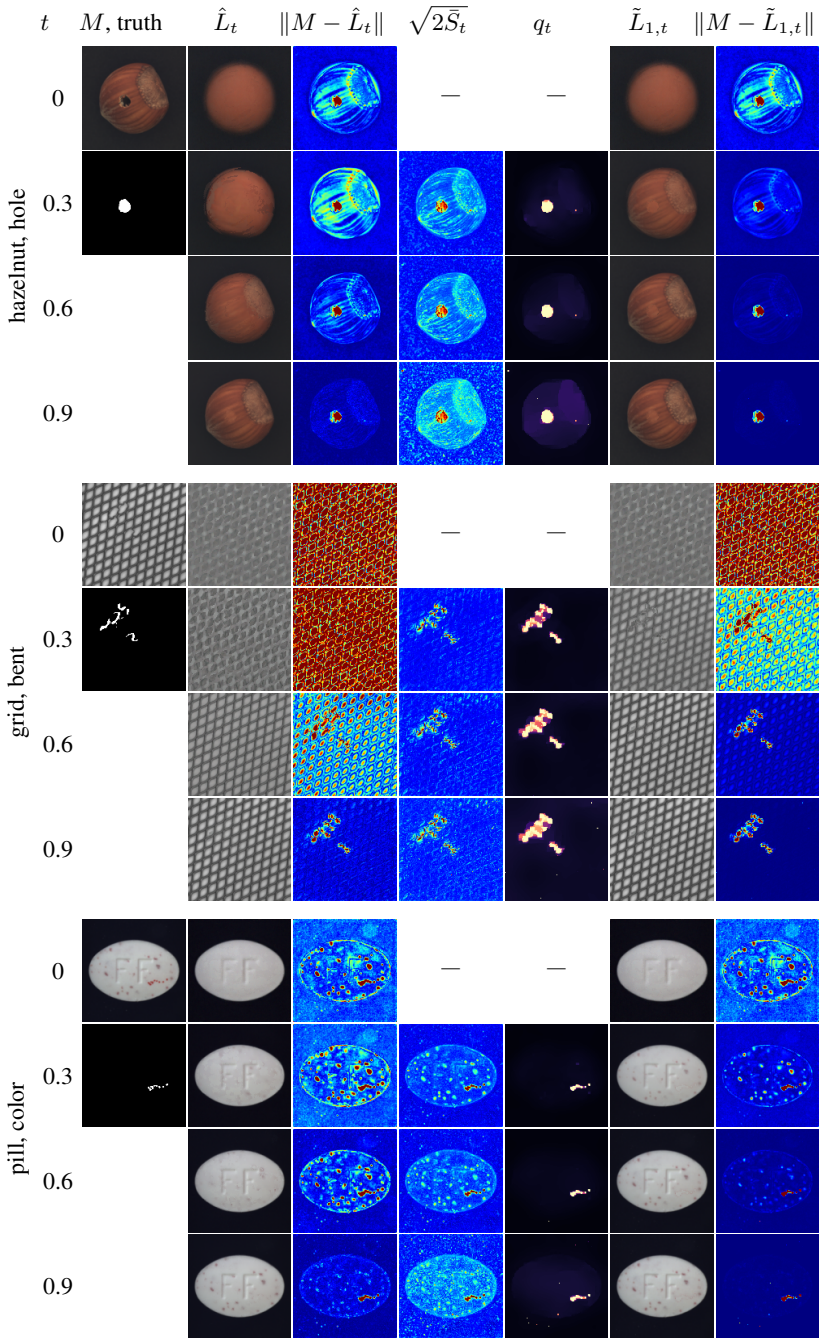


Figure 8: Restoration dynamics of the test images of Figure 2, in the last of the $C = 2$ cycles. Each row is one decision time t of Algorithm 1, and its columns follow one step: the model's endpoint prediction $\hat{L}_t$ on the controlled path and its residual against M; the evidence $\sqrt{2\bar{S}_t}$ accumulated by the probe paths, merged with the first cycle, and the belief q_t solved from it (7); the controller's endpoint $\tilde{L}_{1,t}$ (17) and its residual. At $t = 0$ no measurement has been taken and the state moves by the model alone, so $\tilde{L}_{1,0} = \hat{L}_0$; at $t = 0.9$ the endpoint is the restoration X_1 (20), and its residual is the map A (21) before smoothing. Where the K_c paths differ, each pixel shows the path with the largest residual. Residuals and evidence use the jet colour map from 0 to the 99.5th percentile, over the category's test images, of the unsmoothed A or of $\sqrt{2\bar{S}}$; q uses magma from 0 (dark) to 1 (bright).