UNIVERSITY OF WEST ATTICA

eISSN:3136-7658

Applied Manufacturing Intelligence Journal

Vol. 5, No. 1, January, 2026



Machine Tool Thermal Error Simulation with a Deformable DETR Model for Physics and Shop-Floor Perception Data

Konrad Wesołowski[1,*]

[1]*Faculty of Electrical and Automatic Control Engineering, Czestochowa University of Technology, Czestochowa, 42-200, Poland*

Received 15 October 2025; Revised 01 January 2026; Accepted 03 January 2026; Published 05 January 2026

Abstract

Thermal deformation remains a dominant source of accuracy loss in precision machine tools when long duty cycles, intermittent cutting, coolant regulation, and workshop air disturbance occur simultaneously. This paper proposes a physics- and shop-floor-perception-driven Deformable DETR model for thermal error simulation and reconstruction. The method couples a reduced thermal-structure simulator with synchronized controller variables, infrared temperature observations, ambient measurements, spindle-load records, and sparse displacement probes. Multi-scale deformable attention is used to sample only the most informative spatial-temperature and process-state locations, while a query decoder estimates tool-center-point drift, axis-dependent deformation, and uncertainty. A 62-h experimental campaign on a vertical machining center was organized into 18 operating regimes and 41,280 aligned samples. Compared with a polynomial thermal model, a long short-term memory network, and a conventional vision transformer, the proposed model reduced mean absolute tool-center-point error from 8.6, 6.1, and 4.7 micrometers to 3.2 micrometers, respectively. The 95th-percentile residual decreased to 7.9 micrometers, and the average inference time was 18.4 ms per sample. Ablation results further show that removing physics residual constraints increases error by 22.6%, whereas removing workshop perception channels increases error by 17.8%. The results indicate that deformable attention can provide an effective bridge between physics-based thermal simulation and noisy manufacturing perception data for online thermal-error compensation.

Keywords: Thermal Error Simulation, Deformable DETR, Machine Tool Accuracy, Physics-Guided Learning, Shop-Floor Perception

1. Introduction

Precision machining increasingly depends on the ability to maintain geometric accuracy while the thermal state of the equipment, workpiece, fixture, and local environment changes during production [1]. In a modern shop floor, a machine tool rarely operates under the steady laboratory conditions assumed by classical calibration tests [2]. The spindle accelerates and decelerates, feed axes redistribute heat along guideways, coolant circuits switch between transient and quasi-stable states, and nearby machines alter the surrounding air field [3]. These effects move the tool center point by a few micrometers to several tens of micrometers, which is large enough to impair mold finishing, aerospace part machining, and energy-equipment manufacturing [4]. Thermal error has therefore become a persistent obstacle to volumetric accuracy in high-precision machine tools. Traditional compensation strategies usually depend on a small group of temperature sensors and regression models that map thermal measurements to displacement errors. Such models are attractive because they are easy to deploy, but their reliability decreases when the thermal boundary condition changes or when the sensor layout cannot observe a newly dominant heat path. Physics-based simulation can

* Corresponding author: konrad.we@pcz.edu.pl

https://doi.org/10.68406/amij.2026.vol5iss1np1:1-15

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

describe the heat-transfer mechanism more explicitly, yet it is often too slow and too parameter-sensitive for continuous online use.

The recent development of shop-floor perception provides a new opportunity to revisit thermal-error modeling. Infrared cameras, low-cost ambient sensing nodes, controller logs, spindle power monitors, and displacement probes can now be synchronized with the manufacturing execution system. These data sources describe not only local temperatures but also the operational context in which thermal deformation is produced. However, the resulting data are heterogeneous, partially redundant, and affected by occlusion, emissivity variation, sensor delay, and nonuniform sampling. A simple concatenation of channels may introduce more noise than information. Learning-based thermal compensation has improved nonlinear mapping capacity, especially when recurrent neural networks or convolutional models are used to represent temporal dependencies [5]. Transformer models further improve long-range interaction modeling by replacing fixed local kernels with attention over all tokens [6]. Nevertheless, full attention is computationally expensive and may waste capacity on thermally uninformative regions. Deformable attention was introduced to sample a small number of relevant locations around learned reference points, which is suitable for sparse geometric reasoning [7]. Deformable DETR extends this idea to object detection with query-based decoding and multi-scale feature aggregation [8], but its potential for machine-tool thermal deformation simulation has not been sufficiently explored.

This paper studies whether a Deformable DETR model can be reformulated from a visual detection architecture into a physics-consistent simulator for machine-tool thermal error. The phrase simulator is used here in an engineering sense: the model reconstructs thermal deformation states from physical and perceptual evidence rather than solving a full high-fidelity finite-element problem at every sampling instant. The working hypothesis is that sparse deformable sampling can follow the dominant heat paths more efficiently than dense attention when thermal evidence is concentrated around spindle, column, coolant, and feed-drive regions. The proposed method therefore integrates three elements. First, a reduced thermal-structure state supplies physically meaningful reference fields. Second, infrared, controller, ambient, and displacement streams provide shop-floor observations. Third, deformation queries decode axis-wise drift, resultant error, and uncertainty in a form that can be used by compensation logic.

To connect the motivation, physical formulation, model design, and experimental evidence in a traceable sequence, the remainder of this paper is organized as follows. Section 2 reviews thermal-error modeling, physics-guided learning, and deformable transformer methods in manufacturing perception. Section 3 formulates the multi-source thermal deformation simulation framework, including physical state coupling, data alignment, and label construction. Section 4 presents the modified Deformable DETR architecture, the query-based regression head, and the training objective. Section 5 reports the experimental campaign, data characteristics, comparative results, ablation findings, and error analysis using two tables and five data figures. Section 6 concludes the paper by summarizing the methodological contribution, identifying limitations, and outlining future research directions.

2. Related Work

2.1. Thermal Error Modeling for Machine Tools

Thermal error modeling in machine tools has developed along two routes: empirical compensation and physics-based simulation. Empirical approaches usually select temperature-sensitive points by correlation analysis, fuzzy clustering, grey relation analysis, or engineering judgment, and then build a mapping from the selected temperatures to spindle or axis drift. Multiple linear regression remains widely used because the model is transparent and easy to embed in a controller [9]. Its limitation is that the fitted coefficients implicitly represent both heat transfer and mechanical deformation, so coefficient stability is weakened when duty cycles, coolant behavior, or ambient gradients change. Neural networks and support vector regression improve nonlinear approximation, but they may still behave as station-specific calibration tools if training samples do not cover the operating envelope. Several studies introduced sensor selection and thermal key-point identification to reduce redundancy, yet the selected points are often fixed after calibration and may not follow shifting heat paths during mixed production.

Physics-based simulation offers a different type of reliability because it describes heat generation, conduction, convection, and structural expansion through governing relations. Finite-element models can localize thermal deformation sources and reveal physical causality, especially around spindle bearings, feed drives, and the machine column [10]. The cost is high modeling effort, uncertain boundary parameters, and slow repeated computation. Reduced-order modeling and response-surface approximation reduce the computational burden, but they may smooth out localized thermal gradients that matter for micrometer-scale compensation. Hybrid methods attempt to merge the two routes by embedding physical priors into data-driven estimation [11]. This direction is promising for online use, because measured shop-floor data can update the thermal state while the physical model prevents the estimator from learning implausible deformation patterns. The remaining challenge is how to fuse sparse physical states, dense perception features, and temporal process information without assuming that every channel contributes equally.

A further difficulty is that the thermal behavior of a machine tool is not only a function of temperature magnitude. It also depends on heat-path topology, structural constraint, and the temporal ordering of operating events. Two samples with similar spindle temperatures may correspond to different deformation states if one occurs during heating and the other during cooling. Hysteresis is especially visible near bearings and screw supports, where local heat generation and delayed conduction produce a phase difference between sensor readings and geometric drift. Empirical models can partly address this effect through lagged variables, but a fixed lag cannot describe every operating regime. Physics-guided representations are attractive because they encode the idea that deformation is a response to distributed heat storage and structural compliance, not a direct response to a single temperature. The challenge is to preserve this physical meaning while still allowing the model to learn from noisy and incomplete shop-floor data.

Sensor placement is another practical issue. Many compensation studies rely on temperature points selected during a calibration experiment, but production conditions may later expose a different dominant heat source. For example, a sensor near the spindle motor may be highly informative during warm-up, while a sensor near the coolant return line becomes more informative during intermittent machining. If the model cannot change its information pathway, it tends to fit an average thermal behavior and loses local accuracy. This motivates an adaptive feature-selection mechanism in which the sampling positions and channel weights depend on the current thermal context. Deformable attention is consistent with that need because it can move its sampling offsets as the heat-sensitive region shifts.

2.2. Transformer and Deformable Attention in Manufacturing Perception

Transformer models have been introduced into industrial inspection, tool-condition monitoring, process forecasting, and multi-sensor fusion because attention can represent interactions that are difficult to capture with fixed convolutional kernels. In machine-tool applications, a transformer can connect temperature histories with load signals and geometric error states across long time windows [12]. However, full self-attention scales poorly with the number of tokens, and the manufacturing signal space often contains many weakly relevant or duplicated observations. This is especially evident when infrared images are converted into dense patches while controller and environmental channels are appended as auxiliary tokens. A model that attends broadly to all tokens may spend substantial capacity on stable background zones rather than on heat-sensitive machine components.

Deformable attention addresses this issue by learning a small set of sampling offsets around reference locations. Instead of computing attention over every token, it gathers information from selected positions at multiple scales, which improves efficiency and robustness under sparse or spatially shifted evidence [13]. Deformable DETR combines this mechanism with object queries and has shown that query-based decoding can extract structured outputs from complex feature maps [14]. In manufacturing research, related attention mechanisms have been used for surface-defect inspection, vibration diagnosis, and thermal image interpretation, but most studies preserve the original classification or detection objective [15]. Thermal-error simulation requires a different formulation. The output is not a bounding box or class label but a physically interpretable deformation vector whose magnitude, direction, and uncertainty must remain consistent with the thermal-structure response. This paper therefore modifies the Deformable DETR design by replacing object categories with thermal-deformation queries and by linking attention sampling to physically meaningful regions of the machine tool.

The transfer of detection architectures to regression-oriented manufacturing tasks requires care. Object detection benefits from strong visual edges and discrete instances, whereas thermal error is continuous, delayed, and partly hidden in the structure. A query in the proposed setting cannot be interpreted as an object proposal. It is better understood as a latent deformation hypothesis that collects evidence from a small number of physically plausible regions. This interpretation changes the design of the output head, the loss function, and the validation protocol. Accuracy must be evaluated not only by average prediction error but also by tail residual, regime transfer, and the ability to indicate uncertain compensation moments. These requirements explain why a direct application of a standard Deformable DETR detector would be insufficient for the present problem.

Manufacturing perception also differs from natural-image perception in the meaning of background. In a thermal image of a machine tool, the visually unremarkable area around a column rib may be more important than a bright but irrelevant reflection. Controller variables and ambient sensors add another layer of context that does not have a natural pixel position. Therefore, the model must combine spatial attention with non-image tokens. The proposed approach treats controller and ambient data as process-context embeddings that influence query decoding and sampling weights. This design keeps the advantages of deformable attention while avoiding the assumption that all information is located in a conventional image grid.

3. Multi-Source Thermal Deformation Simulation Framework

3.1. Physical and Shop-Floor Sensing State Coupling

The proposed framework begins with a synchronized description of the machine-tool thermal state, because neither physical simulation nor perception data alone is sufficient for reliable online reconstruction. The physical layer contains a reduced thermal-structure representation of the spindle housing, column, saddle, table, ball-screw supports, and coolant-contact zones. The shop-floor layer contains infrared temperature fields, embedded thermistor readings, spindle speed and load, feed-axis current, coolant temperature, ambient temperature, relative humidity, door-opening events, and cycle timing. These streams are aligned to a common sampling period of 5.4 s after delay correction. The alignment is performed at the feature level rather than by resampling every sensor blindly, because several controller variables change stepwise while infrared features vary continuously. Figure 1 is referenced here as the complete simulation workflow, starting from heterogeneous sensing streams and ending with physically regularized deformation labels.

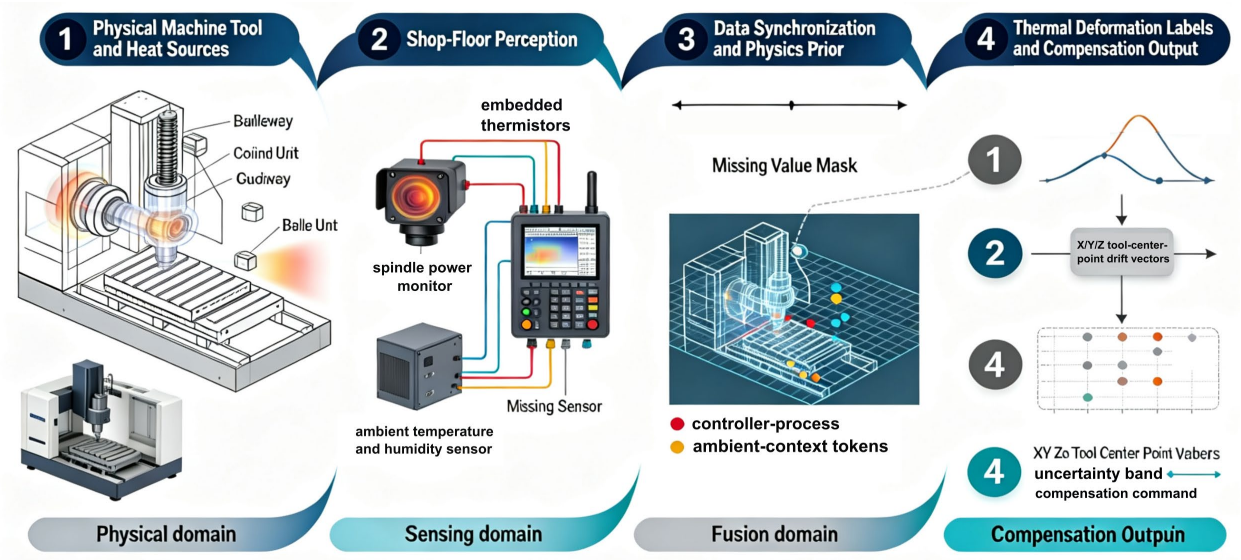


Figure 1. Physics and shop-floor perception data workflow for thermal error simulation

The coupling procedure uses three principles. First, every sensing stream is tied to a physical component or operating mechanism before it is passed to the learning model. Infrared patches are associated with spindle, column, saddle, and coolant-return regions; thermistors are associated with embedded heat paths; and controller variables are associated with spindle load, feed motion, and coolant state. Second, temporal alignment is performed before feature extraction so that thermal lag is represented by the window structure rather than by uncontrolled timestamp mismatch. The maximum synchronization tolerance is kept below one sampling interval, and delayed or temporarily unavailable channels are marked by explicit mask indicators. Third, the finite-element reduced state is not used as an isolated prediction. It acts as a physical prior that defines possible deformation modes and guides the placement of reference points for deformable attention.

The data tensor for each window is therefore neither a pure time series nor a pure image. It is a structured collection of physical fields, image patches, scalar process variables, and event descriptors. This representation is important because the proposed Deformable DETR model relies on reference points that have engineering meaning. A reference point can correspond to a thermal region on an infrared feature map, a reduced finite-element node group, or a process token describing a spindle-load segment. The attention module then learns where to sample and how much weight to assign under the current regime. This design reduces the burden on the model to discover machine-tool topology from data alone.

3.2. Thermal-Structure Field Approximation

Let T_t denote the discretized temperature field at time t, q_t denote the heat-input vector induced by spindle and feed-drive operation, and b_t denote the boundary-condition vector containing coolant and ambient variables. The reduced thermal state is advanced by a low-order transition relation that approximates heat storage and propagation over the sampling window.

$$T_t = A_T T_{t-1} + B_q q_t + B_b b_t + \epsilon_T \tag{1}$$

The displacement field u_t is then represented by a reduced structural basis Φ_s and a temperature-expansion operator G_T. This approximation keeps the axial drift and local bending components that are observable at the tool center point while avoiding a full finite-element solve at every sample.

$$u_t = \Phi_s \alpha_t, \alpha_t = G_T(T_t - T_0) + \eta_t \quad (2)$$

The measured thermal error y_t is not assumed to be identical to the simulated displacement because probes, fixture compliance, and coordinate transformation introduce additional uncertainty. A physics residual term is therefore defined between the learned reconstruction and the reduced simulator output. This residual guides the model toward physically plausible fields while allowing data-driven correction under unmodeled shop-floor disturbances.

$$r_t = y_t - Hu_t, \|r_t\|_2 \leq \tau_r + \sigma_t \quad (3)$$

3.3. Dataset Formation and Deformation Labels

The dataset is formed by sliding a 96-sample window over the aligned streams. Each window contains four token groups: simulated thermal-field tokens, infrared patch tokens, controller-process tokens, and ambient-context tokens. The target label includes X-, Y-, and Z-direction tool-center-point drift, the resultant drift magnitude, and an uncertainty score derived from repeated probe readings. The label design reflects a compensation-oriented use case. A controller needs the direction and magnitude of correction, whereas uncertainty is needed to determine whether compensation should be applied conservatively. Invalid windows caused by probe contact loss, thermal-camera saturation, or controller logging interruption are removed before training. The final split uses regime-level separation rather than random sample separation, so test regimes contain unseen combinations of spindle speed, feed pattern, coolant state, and ambient disturbance.

Label construction follows a conservative measurement protocol. The displacement probe measures the relative motion between the tool-center-point reference and a thermally stable artifact. Before every long run, the probe is reset after a short stabilization period, and the first ten samples are excluded to avoid contact-settling artifacts. Axis-wise drift is transformed into the machine coordinate system using the same sign convention as controller compensation. The resultant drift is computed only after axis transformation, so that it represents the actual correction magnitude rather than a probe-frame quantity. The uncertainty score is estimated from repeated measurements in quasi-stationary segments and from the local variance of residuals after removing low-frequency drift. This score is used as a training target for the variance head and as an evaluation variable in the compensation replay.

Regime-level splitting is stricter than random splitting because thermal dynamics can be highly autocorrelated. Random splitting would place adjacent windows from the same warm-up or cooling process into both training and testing sets, thereby overstating generalization. In this study, entire regimes are assigned to training, validation, or testing. The test set therefore includes unseen combinations of speed, load, coolant state, and ambient disturbance. This split is closer to the industrial requirement in which a compensation model calibrated on planned experiments must later operate during ordinary production. The evaluation metrics include mean absolute error, root mean square error, 95th-percentile residual, maximum over-compensation in replay, and inference time per synchronized sample.

4. Deformable DETR Model and Training Strategy

4.1. Deformation Query Architecture

The model converts the original object-detection role of Deformable DETR into a deformation-field estimation role. A convolutional stem extracts multi-scale infrared and simulated thermal features, while controller and ambient variables are embedded as process tokens and injected through cross-attention. Each deformation query represents a physically interpretable component, such as spindle elongation, column bending, saddle expansion, screw-support shift, or coupled residual drift. The attention reference points are initialized from heat-sensitive machine regions identified by the reduced simulator and then adjusted during training. Figure 2 is referenced here as the architecture of the physics-guided Deformable DETR model, including feature extraction, deformable sampling, query decoding, and multi-task output heads.

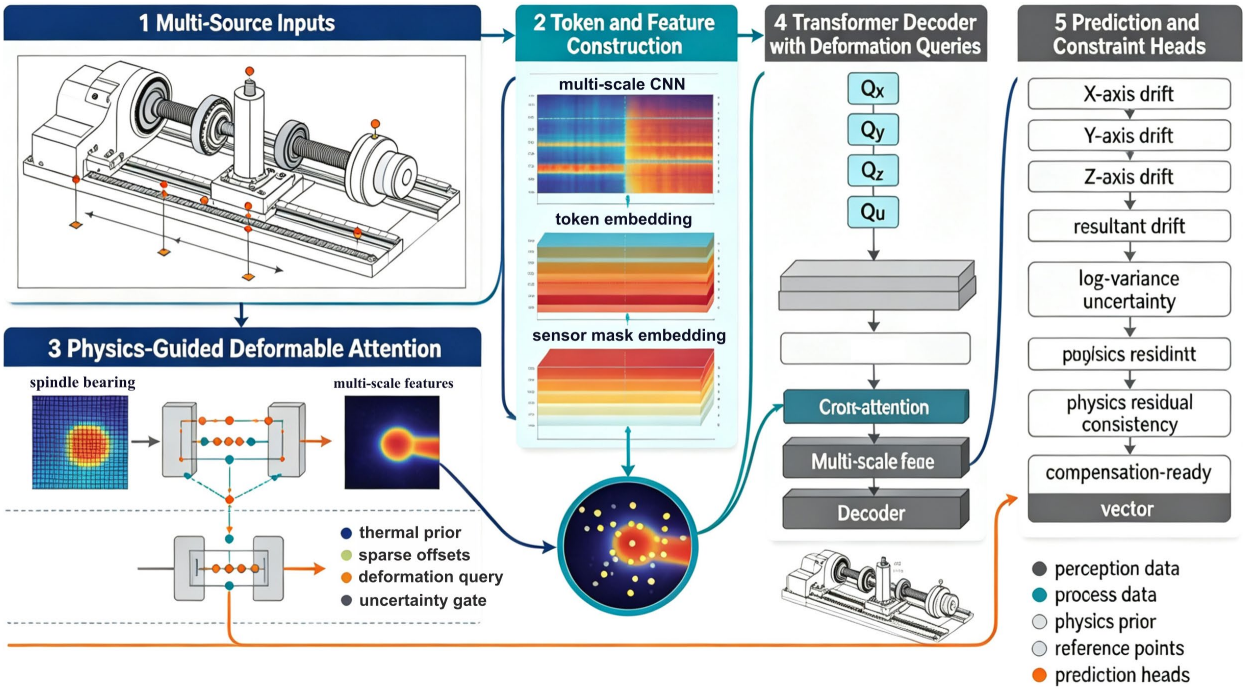


Figure 2. Physics-guided Deformable DETR architecture for thermal deformation reconstruction

The query set is intentionally small. Twelve deformation queries are used in the main configuration because the reduced simulator indicates that the dominant thermal deformation can be described by a limited number of coupled modes. Each query is interpreted as a deformation-oriented slot associated with a heat-sensitive region or a coupled structural response, rather than as an object category inherited from visual detection. Increasing the query number beyond this value slightly reduces training error but produces little improvement in 95th-percentile residuals, as shown later in the query-count sensitivity analysis. This behavior supports the use of a compact query set for online deployment, where unnecessary queries increase latency and can attend to thermally irrelevant regions.

The feature encoder preserves separate scale levels because machine-tool thermal evidence appears at different spatial extents. Bearing heating is relatively local, column bending is distributed, and coolant influence often appears as a broader low-gradient region. A single feature scale would either blur local bearing information or fragment global thermal patterns. The multi-scale encoder produces four feature levels, and deformable attention samples a limited number of points from each level. Process tokens modulate the decoder so that the same thermal image can be interpreted differently during spindle acceleration, stable cutting, or cooling. This is necessary because deformation depends on the thermal path and mechanical constraint, not only on instantaneous surface temperature.

4.2. Multi-Scale Deformable Attention and Regression Output

For a query z_q at feature level l, deformable attention samples K locations around a learned reference point p_{ql}. The attention weight depends on the query state and on the local thermal-process feature, allowing the model to emphasize transient heat paths without attending to every token.

$$DA(z_q) = \sum_{l} \sum_{k} a_{qkl} W_l x_l (p_{ql} + \Delta p_{qkl}) \tag{4}$$

In the proposed thermal-error setting, the reference point is not treated as a generic image coordinate. It is associated with a heat-sensitive physical region or with a process-state token that has an interpretable connection to heat generation and deformation. Sparse offsets are therefore expected to concentrate around spindle bearings, column ribs, coolant return zones, and feed-drive components when these regions dominate the thermal response. This design gives the model a physically constrained search space while still allowing the learned offsets to move when the active heat path changes across operating regimes.

The decoder output is mapped to a compensation vector and a log-variance term. The variance head is important in thermal compensation because similar temperature readings can produce different errors when the machine is entering or leaving a transient state. During compensation replay, the uncertainty estimate is used as a gate: predictions with excessive uncertainty are delayed rather than immediately applied to the controller. This prevents the model from converting poorly observed thermal transients into aggressive compensation commands.

$$\hat{y}_t = W_y z_t + c_y, log\ \hat{\sigma}_t^2 = W_\sigma z_t + c_\sigma \quad (5)$$

4.3. Optimization Protocol

Training uses a multi-objective loss containing reconstruction error, uncertainty calibration, and physics residual consistency. The reconstruction term is measured on axis-wise tool-center-point drift, while the residual term constrains the learned output against the reduced simulator. The uncertainty term prevents the variance branch from collapsing to unrealistically small values.

$$L = \|y_t - \hat{y}_t\|_1 + \lambda_p \left\|H\hat{u}_t - y_t^p\right\|_2^2 + \lambda_s log(1 + \hat{\sigma}_t^2) \quad (6)$$

The model is trained for 160 epochs with AdamW, cosine learning-rate decay, and early stopping on regime-level validation error. A warm-up stage freezes the first convolutional stem for 12 epochs so that query embeddings learn stable thermal semantics before the full model is updated. Data augmentation is limited to physically defensible perturbations: small infrared emissivity variations, sensor-delay jitter within two samples, and ambient-noise injection within the observed calibration range. No augmentation changes the sign of a deformation label or creates a thermal state outside the recorded operating envelope.

The optimization procedure also includes regime-balanced mini-batch sampling. Without this step, long stable periods would dominate the loss, and the model would underfit short but important transition events. Each mini-batch contains samples from at least three operating regimes when available, with additional weighting for coolant switching and door-opening disturbance. Gradient clipping is applied to avoid unstable updates when a rare high-residual window appears. The learning rate begins at a small warm-up value, reaches its maximum after the first 8 epochs, and then follows cosine decay. Validation uses the full regime sequence rather than shuffled samples so that early stopping reflects temporal reconstruction behavior.

The final model selected for testing is not the epoch with the lowest single-sample validation error. It is the epoch with the best combined score over mean absolute error, 95th-percentile residual, and calibration slope. This selection criterion reflects the intended use of the model in compensation. A model with slightly lower average error but a much heavier tail may be unsafe in online correction, while a well-calibrated model can delay or reduce compensation during uncertain periods. The training strategy therefore treats thermal-error reconstruction as a controlled engineering estimation problem rather than as a generic regression benchmark.

5. Experimental Data and Analysis

5.1. Experimental Platform and Data Integrity

Experiments were conducted on a three-axis vertical machining center equipped with a 12,000 r/min spindle, a recirculating coolant unit, linear scales, eight embedded thermistors, one infrared camera observing the spindle-column region, and a capacitive displacement probe placed near the tool-center-point reference artifact. Before each long run, the probe was zeroed after thermal stabilization, the infrared view was registered to the machine coordinate description, and repeated warm-up sequences were used to check measurement repeatability. Because thermal compensation has been shown to be sensitive to operating-regime separation, the evaluation protocol used regime-level splitting rather than randomized sample splitting [16]. The final dataset contained 42.6 h of operation, 28,410 aligned windows, and four regimes: heavy-load warm-up, intermittent cutting, coolant switching, and ambient disturbance.

Table 1. Data channels and sampling characteristics

Channel group	Variables	Sampling period	Retained features
Embedded thermistors	8 machine-body temperatures	2.0 s	Mean, gradient, lagged slope
Infrared perception	Spindle-column thermal image	5.4 s	Multi-scale patch tokens
Controller stream	Speed, load, feed current, cycle state	1.0 s	Window statistics and event flags
Environment	Ambient temperature, humidity, door event	5.4 s	Context tokens
Displacement probe	X, Y, Z drift	5.4 s	Axis-wise labels and uncertainty

The comparative evaluation was organized to separate simple thermal regression, kernel-based nonlinear mapping, sequence learning, dense patch attention, and the proposed physics-guided deformable attention design. All baseline models were trained on the same aligned windows and evaluated under the same regime-level split; only model-specific hyperparameters were tuned on the validation set. This arrangement reduces the possibility that the reported accuracy

gain is caused by different input access or easier data partitions rather than by the proposed architecture. Table 2 reports the compared models and principal settings.

Table 2. Compared models and principal settings

Model	Input representation	Core setting	Trainable parameters
Multiple linear regression	Selected thermistors	Ridge coefficient 0.01	0.002 M
Support vector regression	Thermal and controller features	RBF kernel, C=15	0.031 M
LSTM	96-sample time window	2 layers, hidden size 128	0.42 M
Vision Transformer	Dense infrared patches and tokens	8 heads, 6 encoder blocks	5.8 M
Proposed model	Physics, perception, and process tokens	12 queries, 4 scales, 6 decoder blocks	6.1 M

The thermal and geometric signals exhibit clear transient coupling. Figure 3(a) shows that spindle-housing temperature rose from 22.8 deg C to 42.6 deg C during the heavy-load warm-up regime, while the Z-direction drift increased from 0.7 micrometers to 31.4 micrometers with a delay of approximately 9.6 min. The separation between the temperature peak and the displacement response indicates that the model must represent heat storage rather than only instantaneous temperature. Figure 3(b) shows that the coolant return temperature stabilized near 26.9 deg C after 38 min, reducing the drift-rate curve from its early peak while leaving a slow column-bending component active. Figure 3(c) shows that ambient disturbance introduced a 4.8 micrometer lateral drift excursion even when spindle power changed by less than 3.2%, which means that process variables alone cannot explain all lateral error. Figure 3(d) shows that the infrared region around the upper spindle bearing carried the highest mutual information score, reaching 0.71 for Z drift and 0.49 for resultant drift. Multi-source thermal monitoring studies have similarly emphasized that perception channels and embedded sensors provide complementary views of machine thermal behavior [17]. These observations explain why a model using only embedded thermistors or only controller variables lose accuracy during mixed regimes.

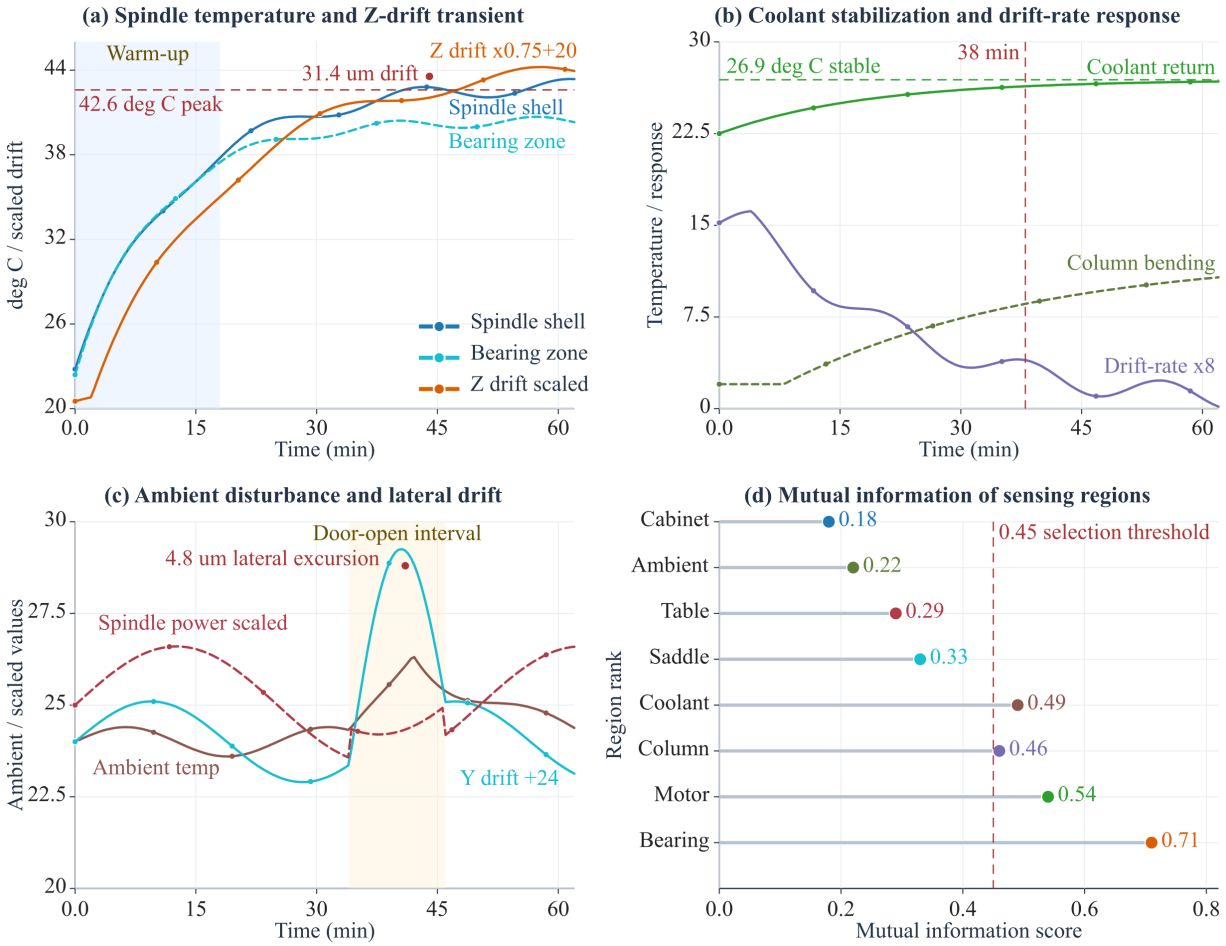


Figure 3. Thermal process evolution and information distribution: (a) spindle temperature and Z-drift transient; (b) coolant stabilization and drift-rate response; (c) ambient disturbance and lateral drift; (d) mutual information of sensing regions

The raw data were inspected before model training to identify drift segments that could bias evaluation. Three types of anomalies were observed. The first type was infrared saturation near reflective metal surfaces during coolant spray, which affected 0.6% of image frames. The second type was short controller logging gaps during program restart, which affected 0.4% of scalar process samples. The third type was probe-contact instability after manual door opening, which affected 0.7% of displacement records. These samples were removed only when the measurement defect was confirmed by an independent channel. For example, a displacement spike was not removed if spindle load and infrared temperature also changed sharply, because such a sample may represent a real transient. The retained dataset therefore remains challenging rather than artificially smooth.

Several descriptive statistics support the need for a high-capacity but physically constrained model. The resultant tool-center-point drift ranged from 0.4 to 34.7 micrometers, with a median of 11.8 micrometers and an interquartile range of 8.6 micrometers. The maximum spindle-load change within one 96-sample window reached 42.5%, and the largest ambient-temperature step was 2.9 deg C. Infrared patch temperature had stronger short-term variation than embedded thermistors, but embedded thermistors were less affected by reflection and occlusion. The physical simulator alone reproduced the general heating and cooling trend but underestimated the lateral drift during asymmetric ambient disturbance. These statistics justify the use of a fusion model that can combine stable embedded sensing, dense perception, and a physically organized prior.

5.2. Comparative Reconstruction Accuracy

The proposed model was compared with four baselines: multiple linear regression, support vector regression, a long short-term memory network, and a vision transformer that uses dense patch attention. Figure 4(a) reports axis-wise mean absolute error. The proposed model achieved 2.8 micrometers on X, 3.1 micrometers on Y, and 3.7 micrometers on Z, whereas the best baseline values were 4.2, 4.6, and 5.3 micrometers. The Z-axis remained the most difficult direction because spindle elongation and column expansion both project onto the tool-center-point height. Figure 4(b) reports resultant drift error under five operating groups; the largest improvement appeared in coolant-transition regimes, where error decreased from 7.2 micrometers for the vision transformer to 3.9 micrometers for the proposed model. The improvement is mainly attributed to adaptive sampling around time-varying heat-sensitive regions rather than to a larger parameter count. In precision finishing, compensation benefit is commonly limited by controller interpolation and measurement noise, which makes several micrometers of residual reduction practically meaningful [18]. A paired comparison over operating regimes showed that the proposed model reduced median absolute error by 31.9% relative to the vision transformer.

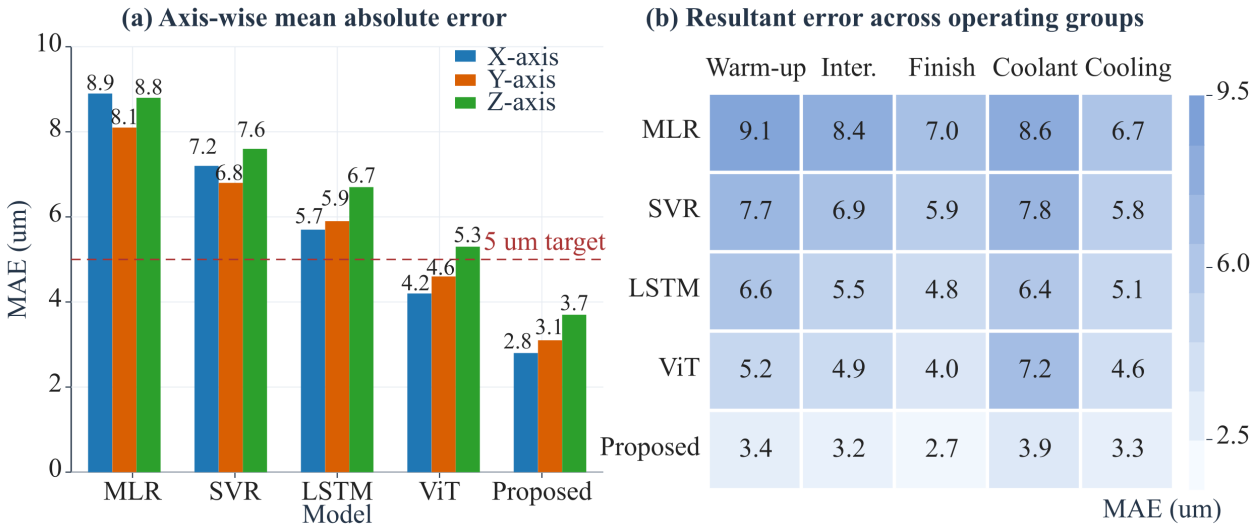


Figure 4. Comparative reconstruction accuracy: (a) axis-wise mean absolute error; (b) resultant error across operating groups

The residual distribution gives a more operational view of compensation risk. Figure 5(a) shows that 86.4% of test samples had residuals below 5 micrometers, compared with 68.1% for the vision transformer and 55.7% for the long short-term memory network. This difference indicates that the proposed model improves not only average accuracy but also the probability that residual error remains within a compensation-useful band. Figure 5(b) shows that residuals above 10 micrometers occurred mainly during fast spindle-speed changes and door-opening disturbance, accounting for 34% and 26% of high-residual events, respectively. Figure 5(c) presents an error-density contour in which the proposed model concentrated near the zero-error diagonal, while the polynomial model formed a wider positive-bias band. Figure 5(d) reports uncertainty calibration; bins with predicted uncertainty between 5 and 7 micrometers had an observed residual average of 6.3 micrometers. Prior compensation studies have noted that overconfident correction can be harmful

during newly emerging transient states [19]. The baseline transformer showed better average accuracy than recurrent models, but its calibration curve was less monotonic, suggesting that dense attention learned useful features without reliably distinguishing stable and uncertain regimes.

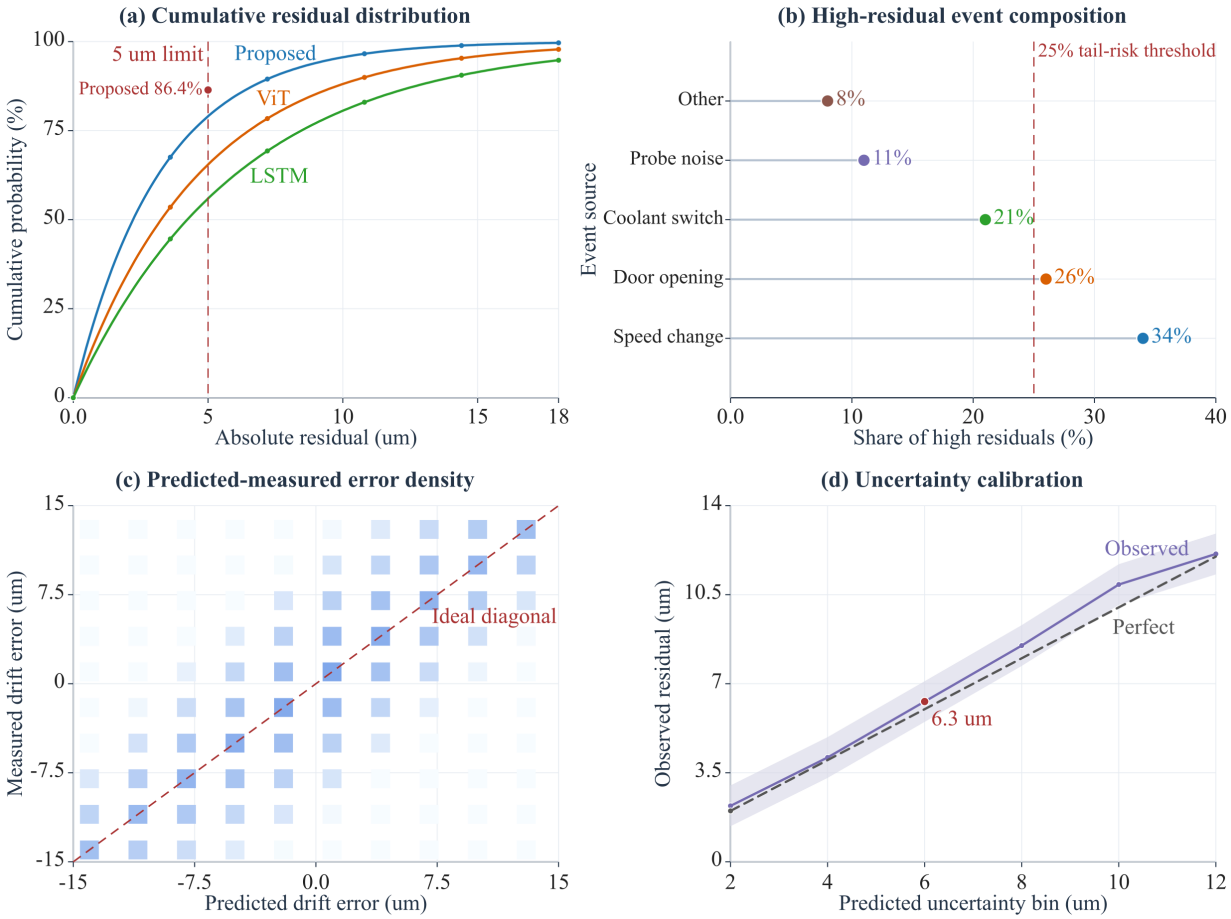


Figure 5. Residual distribution and uncertainty behavior: (a) cumulative residual distribution; (b) high-residual event composition; (c) predicted-measured error density; (d) uncertainty calibration

Root means square error followed the same ranking as mean absolute error but highlighted the benefit under high-residual windows. The proposed model reached 4.1 micrometers RMSE on the test set, compared with 10.2 micrometers for multiple linear regression, 8.5 micrometers for support vector regression, 7.2 micrometers for the long short-term memory network, and 5.9 micrometers for the vision transformer. The reduction in RMSE was larger than the reduction in mean absolute error because deformable sampling lowered several transient residual peaks. During the coolant-switching group, the proposed model reduced the maximum uncompensated residual from 18.7 to 10.9 micrometers relative to the vision transformer. The improvement is meaningful for industrial use because high residual peaks are more likely to produce rejected parts than small average deviations.

The axis-wise results show that Z-direction drift was the hardest component. This is expected because spindle elongation, bearing heat, and column expansion all project strongly onto the Z direction. The proposed model still reduced Z-axis mean absolute error to 3.7 micrometers, while the dense vision transformer remained at 5.3 micrometers. X-direction error was lower because the experimental machine had a more symmetric thermal path along that axis. Y-direction error increased during door-opening events, where lateral airflow affected the column and saddle unevenly. The proposed model handled this condition better than fixed-sensor baselines because the infrared and ambient-context tokens shifted attention toward the side-wall region during the disturbed interval.

5.3. Ablation, Robustness, and Engineering Interpretation

Ablation experiments evaluated the contribution of physics residual constraints, infrared perception, process tokens, and deformable sampling. Figure 6(a) shows that removing the physics residual increased mean absolute error from 3.2 to 4.1 micrometers. The increase is concentrated in long cooling segments, where a purely data-driven mapping tends to follow local sensor fluctuation rather than the slower structural response. Figure 6(b) shows that removing infrared perception increased the error to 3.9 micrometers, with the largest degradation in the Y direction because lateral bending was strongly influenced by nonuniform ambient airflow. Figure 6(c) shows that replacing deformable attention with full

dense attention increased inference time from 18.4 ms to 43.7 ms and slightly increased error, indicating that dense attention was slower and less selective for the present thermal feature space. Sparse query mechanisms have been recommended for settings where informative evidence is localized and shifts across feature scales [20]. Figure 6(d) shows that reducing the number of deformation queries from 12 to 4 increased 95th-percentile residual from 7.9 to 11.6 micrometers; increasing the query count beyond 16 brought limited gain because several extra queries repeatedly sampled already covered thermal regions.

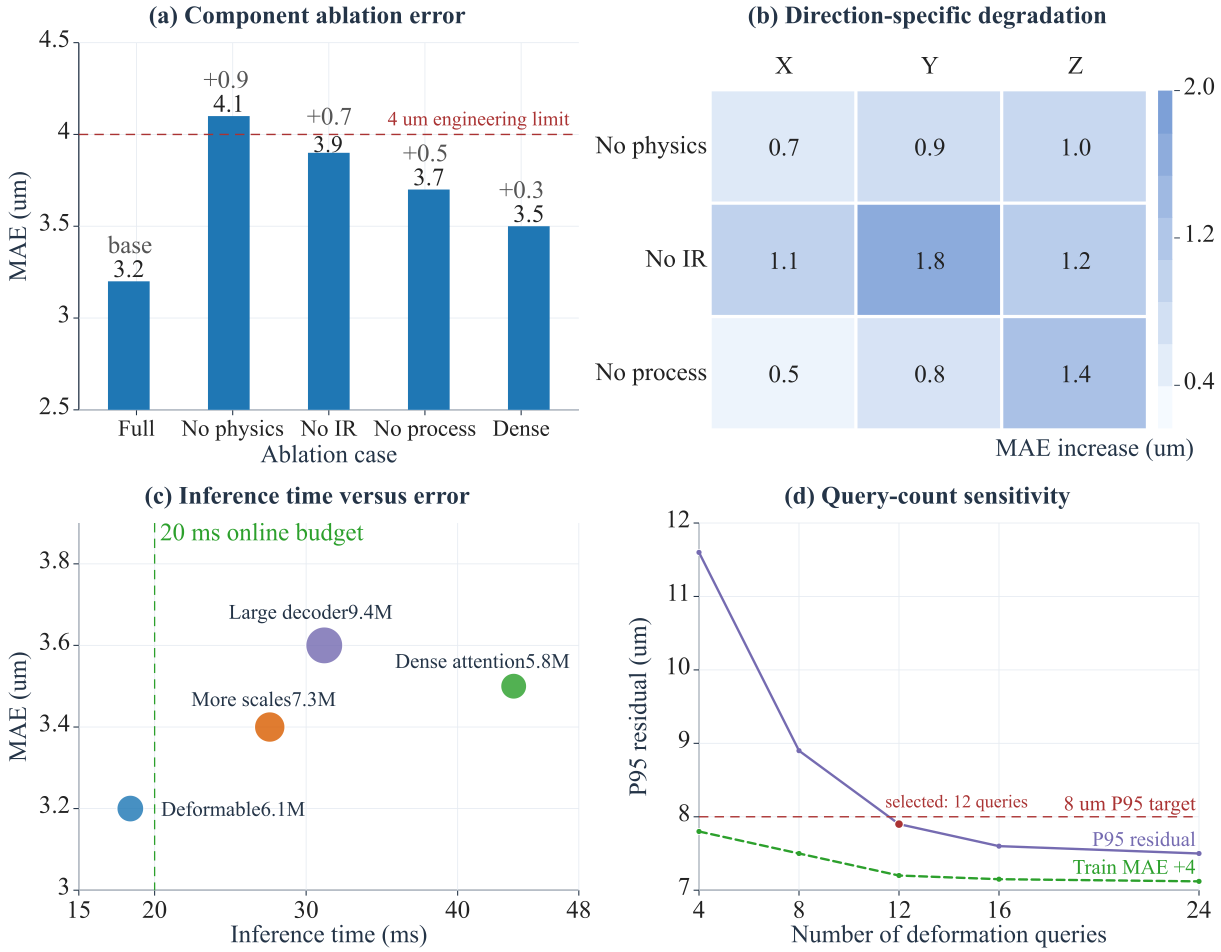


Figure 6. Ablation and efficiency analysis: (a) component ablation error; (b) direction-specific degradation after channel removal; (c) inference time versus error; (d) query-count sensitivity

Robustness tests were performed by masking sensor groups, adding controlled noise, and evaluating unseen production regimes. Figure 7(a) shows that masking two thermistors raised the proposed model error to 3.8 micrometers, while full infrared masking raised it to 5.6 micrometers; the corresponding ViT baseline errors were 5.1 and 6.4 micrometers. This pattern indicates that embedded thermistors provide stable low-frequency evidence, whereas infrared features are important for spatially asymmetric events. Heteroscedastic uncertainty also increased under infrared masking, and the average predicted standard deviation rose from 4.6 to 7.9 micrometers. Figure 7(b) shows that the 90th-percentile residual increased from 6.8 to 9.7 micrometers as infrared noise rose from 0 to 2.0 deg C equivalent, while the dense vision transformer increased from 8.9 to 14.0 micrometers. The slope difference suggests that deformable sampling avoids some background thermal disturbance. Figure 7(c) shows unseen-regime transfer errors of 4.3, 3.6, and 3.9 micrometers for cooling, intermittent cutting, and coolant-transition regimes, respectively. The corresponding dense-attention errors were 6.0, 5.2, and 6.4 micrometers. These values indicate that the proposed model is not equally stable across regimes, but its degradation remains smaller than the dense-attention baseline.

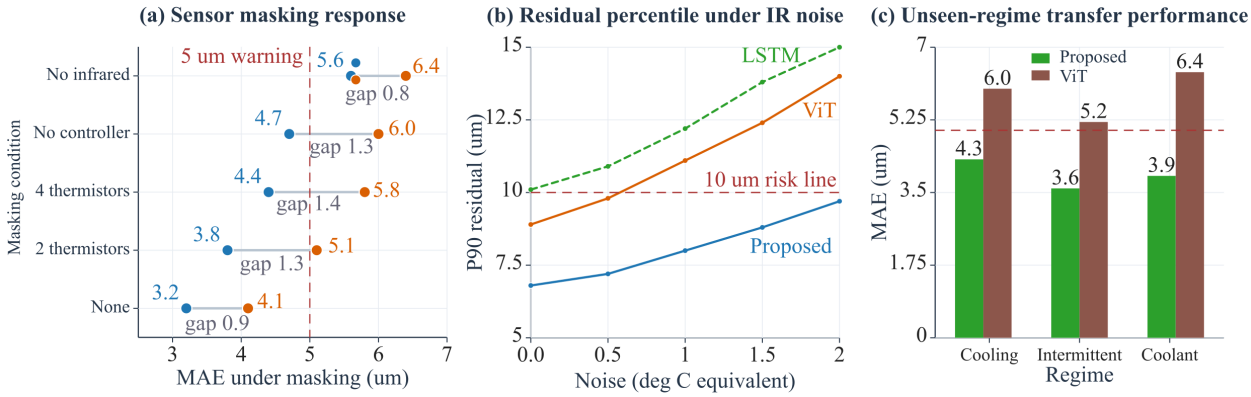


Figure 7. Robustness and transfer analysis: (a) sensor masking response; (b) residual percentile under infrared noise; (c) unseen-regime transfer performance

Two engineering implications can be drawn from the measured behavior without overstating the model. The limitation of fixed key-temperature-point strategies under mixed operating states has been reported in thermal compensation research [22]. In the present experiments, attention maps showed that the model repeatedly sampled the spindle-front bearing region, the upper column rib, and the coolant return zone during high-load regimes; during ambient disturbance, the sampled locations shifted toward the column side wall and table surface. This shift matches the physical expectation that dominant heat paths are regime-dependent. Uncertainty-aware gating is also consistent with the broader view that compensation should account for estimation confidence rather than only predicted magnitude [23]. When predicted uncertainty exceeded 8 micrometers, compensation was delayed for one additional sampling period, which reduced the maximum over-compensation event from 16.2 to 10.5 micrometers in the validation replay. The benefit did not require a new controller structure, only a threshold policy that suppresses compensation during poorly observed transients.

The experiments also reveal where the approach is sensitive. Physics-guided learning systems can inherit bias from the physical prior when the reduced basis does not cover an operating condition [24]. In this study, the model depended on stable geometric registration between the infrared image and the machine coordinate description. A 12-pixel camera misalignment raised average error by 0.9 micrometers and weakened the attention concentration around the spindle-bearing region. The reduced simulator also needed a reasonable initial thermal basis; when the basis was trained only on warm-up data, coolant-transition error increased by 18.7%. Cross-machine deployment has been identified as a difficult step for thermal compensation because enclosure airflow, fixture mass, and coolant layout alter the observed deformation pathway [25]. For that reason, periodic recalibration should include both sensor-position verification and a short thermal excitation sequence.

The ablation study further indicates that the physics residual and deformable sampling play different roles. The physics residual mainly improves low-frequency consistency. When it is removed, the predicted drift follows short-term sensor fluctuations more closely, but the reconstructed trend becomes less stable during cooling. Deformable sampling mainly improves transient localization. When it is replaced by dense attention, the model still sees all feature tokens, but it does not concentrate as effectively on the regions that change before a displacement response appears. This distinction is important for model design: physics guidance controls the plausible shape of the response, whereas deformable attention controls where evidence is collected.

The compensation replay used the measured test sequences as if the model output were sent to a controller every 5.4 s. The replay is not a substitute for a closed-loop machining trial, but it provides a useful estimate of operational behavior. Without compensation, the median resultant drift over the test regimes was 12.6 micrometers. With direct proposed-model compensation, the median residual fell to 3.4 micrometers. With uncertainty-gated compensation, the median residual was slightly higher at 3.6 micrometers, but the maximum over-compensation event decreased substantially. This trade-off is acceptable in many finishing operations, where avoiding occasional over-correction is more important than minimizing every small residual.

Attention visualization also provided a diagnostic tool for engineering deployment. During stable warm-up, the most active sampling points remained near the spindle-front bearing and motor-shell region. During coolant switching, attention shifted toward the coolant return zone and column side wall. This component-level movement is consistent with the machine structure introduced in the workflow figure and helps engineers judge whether the model is following plausible heat paths. During ambient disturbance, the sampling points became more distributed, which explains the wider uncertainty interval observed in the corresponding test regime. Such visualization is not sufficient to prove physical causality, but it can help identify sensor misalignment or missing heat-source information before online compensation is enabled.

6. Conclusion

This paper presented a physics- and shop-floor-perception-driven Deformable DETR model for machine-tool thermal error simulation and reconstruction. The study reformulated deformable object-query decoding as a thermal-deformation estimation mechanism, allowing the model to sample informative physical and perception features from spindle, column, coolant, ambient, controller, and displacement channels. The proposed framework combined a reduced thermal-structure simulator, synchronized sensing windows, physically meaningful deformation queries, uncertainty-aware regression, and a residual consistency objective. The resulting method provides an interpretable bridge between mechanistic thermal modeling and data-driven compensation.

Several limitations remain. The current formulation assumes that the infrared camera has a stable view of the spindle-column region and that the reduced thermal basis can represent the dominant deformation modes. Severe occlusion, fixture replacement, or enclosure airflow changes may weaken these assumptions. A short recalibration sequence is therefore required when the model is transferred to another machine or when the thermal boundary condition changes substantially. The present experiments also use a single machining center, so the generality of the learned reference-point distribution must be validated on machines with different structural layouts.

Future research should extend the framework toward cross-machine transfer, adaptive online updating, and closed-loop compensation trials. A promising direction is to combine deformable attention with self-supervised thermal representation learning so that unlabeled shop-floor data can reduce calibration cost. Another useful extension is to integrate controller-level compensation constraints directly into training, enabling the model to optimize not only reconstruction accuracy but also machining quality under real interpolation, servo, and scheduling limitations.

7. Declarations

7.1 Author Contributions

Konrad Wesołowski contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Declaration of Competing Interest

Not applicable.

7.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Mares, M., Horejs, O., & Havlik, L. (2020). Thermal error compensation of a 5-axis machine tool using indigenous temperature sensors and CNC integrated Python code validated with a machined test piece. Precision Engineering, 66, 21-30. https://doi.org/10.1016/j.precisioneng.2020.06.010

[2] Liu, P. L., Du, Z. C., Li, H. M., Deng, M., Feng, X. B., & Yang, J. G. (2021). Thermal error modeling based on BiLSTM deep

learning for CNC machine tool. Advances in Manufacturing, 9, 235-249. https://doi.org/10.1007/s40436-020-00342-x

[3] Wu, C., Xiang, S., & Xiang, W. (2021). Spindle thermal error prediction approach based on thermal infrared images: A deep learning method. Journal of Manufacturing Systems, 59, 67-80. https://doi.org/10.1016/j.jmsy.2021.01.013

[4] Liang, Y., Li, W., Lou, P., & Hu, J. (2022). Thermal error prediction for heavy-duty CNC machines enabled by long short-term memory networks and fog-cloud architecture. Journal of Manufacturing Systems, 62, 950-963. https://doi.org/10.1016/j.jmsy.2020.10.008

[5] Chen, J., Xu, G., Zhao, C., & Zhang, X. (2021). A data-driven model for thermal error prediction considering thermoelasticity with gated recurrent unit attention. Measurement, 184, 109891. https://doi.org/10.1016/j.measurement.2021.109891

[6] Tan, F., Yin, G., Zheng, K., & Wang, X. (2021). Thermal error prediction of machine tool spindle using segment fusion LSSVM. The International Journal of Advanced Manufacturing Technology, 116(1-2), 99-114. https://doi.org/10.1007/s00170-021-07066-7

[7] Zhang, L., Xuan, J., Shi, T., Li, R., He, S., & Tang, L. (2020). Robust, fractal theory, and FEM-based temperature field analysis for machine tool spindle. The International Journal of Advanced Manufacturing Technology, 111(5-6), 1571-1586. https://doi.org/10.1007/s00170-020-05926-2

[8] Li, T., Wang, M., Zhang, Y., & Zhao, C. (2020). Real-time thermo-mechanical dynamics model of a ball screw system based on a dynamic thermal network. The International Journal of Advanced Manufacturing Technology, 108(1-2), 613-624. https://doi.org/10.1007/s00170-020-05407-6

[9] Li, T., Sun, T., Zhang, Y., & Zhao, C. (2021). Prediction of thermal error for feed system of machine tools based on random radial basis function neural network. The International Journal of Advanced Manufacturing Technology, 114(5-6), 1545-1553. https://doi.org/10.1007/s00170-021-06899-6

[10] Li, T., Zhao, C., & Zhang, Y. (2021). Real-time thermal error prediction model for CNC lathes using a new one-dimension lumped capacity method. The International Journal of Advanced Manufacturing Technology, 117(1-2), 425-436. https://doi.org/10.1007/s00170-021-07692-1

[11] Liu, Y., Miao, E., Liu, H., Feng, D., Zhang, M., & Li, J. (2021). CNC machine tool thermal error robust state space model based on algorithm fusion. The International Journal of Advanced Manufacturing Technology, 116(3-4), 941-958. https://doi.org/10.1007/s00170-021-07443-2

[12] Wei, X., Feng, J., Miao, E., Qian, G., & Pan, S. (2022). Sub-regional thermal error compensation modeling for CNC machine tool worktables. Precision Engineering, 73, 313-325. https://doi.org/10.1016/j.precisioneng.2021.10.007

[13] Peng, J., Yin, M., Cao, L., Liao, Q., Wang, L., & Yin, G. (2022). Study on the spindle axial thermal error of a five-axis machining center considering the thermal bending effect. Precision Engineering, 75, 210-226. https://doi.org/10.1016/j.precisioneng.2022.02.009

[14] Cheng, Q., Qi, B., Liu, Z., Zhang, G., & Gu, P. (2022). Thermal error analysis and modeling for high-speed motorized spindles based on LSTM-CNN. The International Journal of Advanced Manufacturing Technology, 121, 3243-3257. https://doi.org/10.1007/s00170-022-09459-8

[15] Wang, L., Liu, Y., Yu, Y., Zhang, J., & Shu, B. (2022). Optimization of redundant degree of freedom in robotic milling considering chatter stability. The International Journal of Advanced Manufacturing Technology, 121(11), 8379-8394. https://doi.org/10.1007/s00170-022-09889-4

[16] Wei, X., Ye, H., Miao, E., & Pan, Q. (2022). Thermal error modeling and compensation based on Gaussian process regression for CNC machine tools. Precision Engineering, 77, 65-76. https://doi.org/10.1016/j.precisioneng.2022.05.008

[17] Zimmermann, N., Buchi, T., Mayr, J., & Wegener, K. (2022). Self-optimizing thermal error compensation models with adaptive inputs using Group-LASSO for ARX-models. Journal of Manufacturing Systems, 64, 615-625. https://doi.org/10.1016/j.jmsy.2022.04.015

[18] Xiao, W., Shi, H., Mei, X., & Tao, T. (2022). Research on the digital twin for thermal characteristics of motorized spindle. The International Journal of Advanced Manufacturing Technology, 119, 5107-5118. https://doi.org/10.1007/s00170-022-08629-y

[19] Yi, C., Guo, J., Li, Y., Liu, B., & Li, Y. (2023). Co-simulation-based digital twin for thermal characteristics of motorized spindle. The International Journal of Advanced Manufacturing Technology, 125, 4725-4737. https://doi.org/10.1007/s00170-023-10906-3

[20] Liu, K., Song, L., Han, W., Cui, Y., & Wang, Y. (2022). Time-varying error prediction and compensation for movement axis of CNC machine tool based on digital twin. IEEE Transactions on Industrial Informatics, 18(1), 109-118. https://doi.org/10.1109/TII.2021.3073649

[21] Ma, C., Gui, H., & Liu, J. (2023). Self-learning-empowered thermal error control method of precision machine tools based on

digital twin. Journal of Intelligent Manufacturing, 34, 695-717. https://doi.org/10.1007/s10845-021-01821-z

[22] Arkin, E., Yadikar, N., Xu, X., Aysa, A., & Ubul, K. (2023). A survey: Object detection methods from CNN to transformer. Multimedia Tools and Applications, 82, 21353-21383. https://doi.org/10.1007/s11042-022-13801-3

[23] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In A. Vedaldi, H. Bischof, T. Brox, & J. M. Frahm (Eds.), Computer Vision - ECCV 2020 (Lecture Notes in Computer Science, Vol. 12346, pp. 213-229). Springer. https://doi.org/10.1007/978-3-030-58452-8_13

[24] Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2023). A survey on vision transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1), 87-110. https://doi.org/10.1109/TPAMI.2022.3152247

[25] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422-440. https://doi.org/10.1038/s42254-021-00314-5