Journal of Applied Automation Technologies

Research Paper

Boundary-Refined DeepLabV3+ for Short- and Medium-Term Load Forecasting in Regional Power Grids

Lena Majewska[1] and Beata Kubicka[1, *]

[1] Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Krakow, 30-059 Krakow, Poland
*Corresponding author: beata.k@agh.edu.pl

Abstract. Accurately predicting the electricity demand in a specific area is getting harder and harder because of changes in weather, coordinated demand response and all sorts of urban activities at different times and places. Develop a boundary-refined DeepLabV3+ model in this paper to recast the multi-regional load prediction problem as dense regression of a time-region load map. The encoder uses dilated multi-scale context and employs a particularly refined branch to study slope initiation, peak shoulder, valley bottom recovery, and inter-regional transition zones. By optimizing boundary-weighted regression, gradient consistency, and regional consistency loss, the conditional decoder can provide predictions ranging from 24 to 168 hours. The three years of hourly observations from the 24 regional nodes in the experiment have generated 630,720 load records, as well as synchronous meteorological and calendar variables. On the 24h task, the proposed model has a mean absolute percentage error of 1.86% and a root mean square error of 41.7 MW, and these errors are lower than those of the best baseline by 18.4% and 15.2%, respectively. At 168 h, the mean absolute percentage error is 2.94%. The peak-timing deviation is 1.42 h; under missing-weather and heat-wave tests, the performance degradation is only 0.31 and 0.44 percentage points, respectively. According to the above findings, explicit boundary learning can not only improve overall accuracy but also address specific issues such as ramp-up location, regional consistency, and robustness. Therefore, in the practical application of day-ahead and week-ahead scheduling for regional power grids, explicit boundary learning is of significant importance.

Keywords: *DeepLabV3+, Load Forecasting, Spatiotemporal Modeling, Multiscale Features, Energy Management*

Received on 26 June 2023, Accepted on 17 November 2023, Published on 25 November 2023

Introduction

Regional power grids must balance demand at a resolution fine enough for unit commitment, reserve procurement, inter-area transfer, and maintenance scheduling. Electrification of transport and heating has increased the amplitude and speed of local ramps, while distributed resources can conceal demand until coordinated charging or weather-sensitive consumption suddenly emerges [1]. System-level forecasts therefore no longer describe the stresses seen by individual zones [2]. Day-ahead decisions need accurate peak magnitude and timing, whereas week-ahead planning needs stable trends across working days, weekends, and weather regimes [3]. A relatively small overall forecasting error may still place an incorrect reserve in the wrong area [4].

Classical regression and autoregressive models are still relatively interpretable but cannot handle the non-linear cross-effects among temperature, humidity, calendar state, and lagged demand [5]. Recurrent networks have extended the scope of temporal memory, but sequential computation and error accumulation may be difficult for sudden changes [6]. Convolutional and attention models are more general, and conventional objectives distribute errors evenly, thus paying less attention to narrow ramps that are significant for operators [7]. Recent space-time methods use regional coupling but many still assume a fixed graph or reduce all nodes to a single pooled representation [8].

Treat the regional forecast as a dense load-map regression problem and develop a boundary-refined DeepLabV3+ model. Hours are on the horizontal axis, regional nodes are on the vertical axis, and synchronous external factors are represented as channels. Atrous Context considers periodic and cross-regional structures, and the high-resolution boundary branches retain the slope initiation, peak shoulder, valley bottom recovery, and transition zones. Joint regression, gradient, boundary and coherence objectives all support 24h, 72h and 168h forecasting without modifying the core model. Engineering formulation connects dense prediction to operational load transitions, a refinement architecture that recovers details lost during encoding, and all-round evaluation of accuracy, robustness, transfer, event localisation and computational cost.

Related Work

Short- and Medium-Term Load Forecasting

Given the multi-seasonality and non-linear weather response of the regional demand data, deep temporal models are now frequently used instead of shallow learners for data-driven load forecasting. Hybrid convolutional-recurrent models learn local motifs and then propagate memory to enhance household or feeder predictions [9]. Theory-guided recurrent ensembles can add load ratios or physical priors to improve robustness against noisy weather forecasts [10]. Transformers further increase the receptive field and support parallel training, but their peak estimates are still sensitive to a limited number of events [11]. Graph and clustering frameworks show the connections between buses or areas and improve the space-coherence of the assumption of adjacency [12].

The practical deficiency is not an average error rate. The start and end times of the steep ramp, the time of the system peak, and the area of regional stress are needed by dispatchers. Most models reduce pointwise loss, so even if the shoulders of a wide plateau are offset, they still receive relatively high scores. Long horizons increase this weakness because recursive or autoregressive decoders smooth high-frequency structures. Therefore, a model that allocates capacity to transitional communities is reasonable.

Dense Prediction and Boundary Refinement

To preserve multi-scale context, DeepLab's dense predictions combine an encoder and atrous spatial pyramid pooling, and use a decoder to recover high-resolution details [13]. This structure provides useful inductive bias for the time-space matrix when used beyond traditional vision. The dilation spans intraday and weekly cycles without the need for repeated sampling. Boundary-aware semantic segmentation research shows that explicit edge losses can correct uncertain contours [14]. Multiple-loss constraints and residual cascades have also improved DeepLabV3+ boundary quality [15].

Directly importing a semantic segmentation network is nevertheless insufficient. Load values are continuous, regions have unequal scales, and a temporal boundary is an operational transition rather than an object contour. Load maps also require causal inputs and horizon-aware decoding. Existing work rarely connects boundary confidence with load-gradient magnitude, regional consistency, and forecast calibration in one model [16]. The present design adapts boundary refinement to continuous load fields and evaluates whether this adaptation improves event localization as well as conventional errors [17].

Boundary-Refined DeepLabV3+ Model

Spatiotemporal Load-Map Construction

At each time step, the input cube contains the previous 168 hours of load data for the 24 regions, temperature, humidity, wind speed, calendar embeddings and a weather forecast channel. Load is classified by area of training means and standard deviations, and meteorological factors are classified according to the world's standards. Regions are organized based on the proximity of power and historical relevance; therefore, closely connected areas may be near each other, but they do not form a fixed pattern. The output is a dense matrix for the selected 24-, 72- or 168-h horizon.

$$X_t = [L_{t-167:t}, W_{t-167:t+H}, C_{t-167:t+H}] \quad \text{Eq.(1)}$$

The historical and known-future channels are separated by a causal mask here. The target matrix includes time ramps and cross-regional transition zones, and contains the standardized future demand for each region.

$$Y_t(h,r) = \frac{L_{t+h,r} - \mu_r}{\sigma_r} \quad \text{Eq.(2)}$$

A soft boundary target is obtained from temporal and regional gradients. The threshold is estimated from the 85th percentile of training gradients, then softened by a logistic function so near-threshold transitions still contribute.

$$G_t(h,r) = \sqrt{(\Delta_h Y_t)^2 + \lambda_r(\Delta_r Y_t)^2} \quad \text{Eq.(3)}$$

The boundary probability combines magnitude and threshold rather than using brittle binary labels.

$$B_t(h,r) = sigmoid\left(\frac{G_t(h,r) - \tau_g}{s_g}\right) \quad \text{Eq.(4)}$$

Window construction is intentionally anchored to the time of the forecast problem rather than to calendar-day files. The above selection does not allow a minor leak path for the new weather data or completed daily statistics to reach the model prematurely. For each training example, only the meteorological forecast vintage that would have been available at the time of issue is stored. Calendar channels encode hour, day of the week, month, proximity to holiday and working-day status using cyclic sine-cosine pairs and learned categorical embeddings. Cyclic encoding avoids the artificial discontinuity at 23:00-00:00, and embedding separately treats nominally similar days with different activity patterns. The missingness mask is added next to each imputed channel. Therefore, a filled value is used to restore numerical continuity without falsely assuming that the original observation is correct.

The regional order is learned only once from the training partition. The first is an electrical-distance index; among other nodes in the same network layer, a load-correlation index will be selected as a tie-breaker. The dilated encoder discovered a broad neighborhood, while the consistency objective is based on the original adjacency matrix; the sensitivity test with ten random perturbation orders only changed the 24-hour MAPE by 0.06 percentage points. Order still improves convergence: validation loss reaches a lower value seven epochs sooner than in the random layout. Thus, the representation should be seen as a computational atlas rather than a statement that close proximity to the land does not imply electrical contact. Each pixel has a traceable hour and area, and the error map can be returned directly to the operational user.

Boundary-Refined DeepLabV3+

An encoder with a modified ResNet backbone, featuring an output stride of 16. The dilation rates of the branches are 1, 6, 12, and 18, involving local slopes, daily structures, and multi-day envelopes. A global pooling branch supplies system context. Low-level features at output stride 4 are projected to 48 channels and then fused with the upsampled context. A parallel boundary stream receives shallow gradients and intermediate encoder features to produce a gate for modulating decoder activations.

As shown in Figure 1, all the above operations are aimed at prediction: the Atrous branch handles different repetition cycles, the low-level skip module maintains a narrow transition area, and the boundary gate expands the slope neighborhood before the final regression.

$$F_{aspp} = Concat_{d\in\{1,6,12,18\}}\left(Conv_{3\times3}^{d}(F_e)\right) \quad \text{Eq.(5)}$$

Boundary-gated fusion applies a residual modulation so uncertain edges cannot suppress the contextual representation completely.

$$F_g = F_d \odot \left(1 + \alpha sigmoid(F_b)\right) \quad \text{Eq.(6)}$$

The forecast head is conditioned on a learned horizon embedding. This mechanism shares the encoder while allowing the decoder to adapt its smoothing strength to the requested horizon.

$$\hat{Y}^H = \phi_H\left(Up\left(F_g\right) + E_H\right) \quad \text{Eq.(7)}$$

Choosing the encoder width is to preserve the detailed regional structure and avoid visual parameterization of the dataset. A 3x3 convolution is used instead of a 7x7 kernel for the stem, and the first pooling operation is omitted. Modify to keep only narrow one-hour transitions visible at a shallow depth. Residual stages have 64,128,256 and 512 channels, followed by a 256-channel atrous module. Due to the batch size limiting the threedimensional cube input, group normalization is used instead of batch normalization. Contextual features are connected to the low-level features of the projection through bilinear scaling, and then processed by two depthwise separable convolutions. Dropout is applied after fusion and not in the boundary stream; otherwise, random removal of edge evidence resulted in an unstable peak timing during the initial tests.

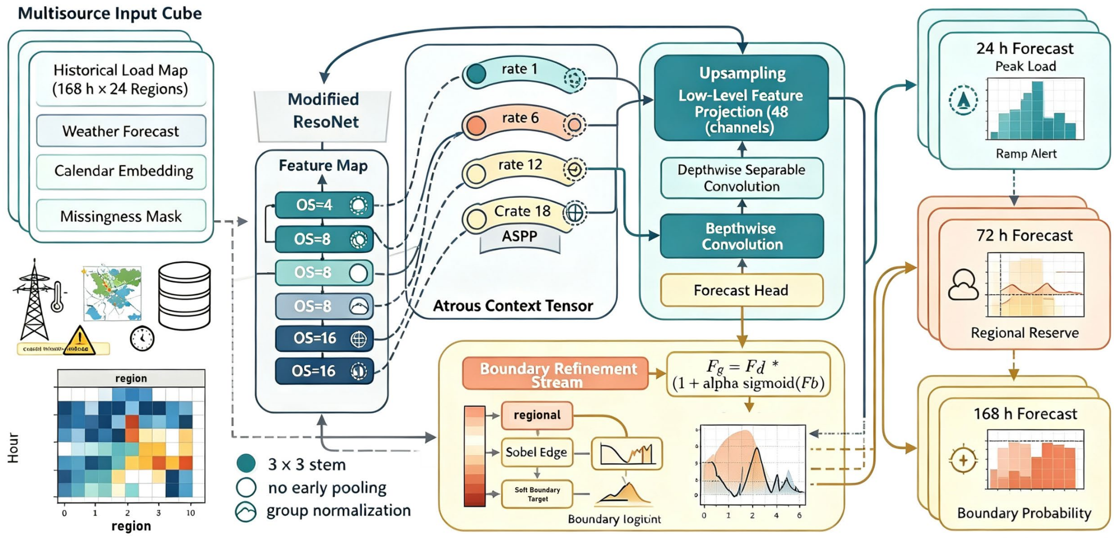


Figure 1. Boundary-refined DeepLabV3+ architecture for dense regional load-map forecasting, including causal input channels, atrous context, high-resolution boundary stream, gated fusion, and multi-horizon output heads.

The refinement stream has two sources. A shallow path receives temporal and regional finite differences of the normalised load, and a semantic path receives output-stride-eight encoder features. Together, they can help the branch differentiate a real system ramp from a single corrupted meter. Boundary logits are upsampled to the output resolution and used twice: once as an auxiliary prediction supervised by soft labels, and again as a bounded gain for the regression decoder. The gain coefficient is set to zero initially; thus, the first few training rounds operate as a standard DeepLabV3+ model. With a stable boundary loss, the coefficient will approach its learned value of 0.43. This warm start was more stable than using a fixed edge gain for the first batch.

Boundary and Context Coordination

Raw edge supervision can overreact to measurement noise. The refinement branch therefore combines gradient magnitude with a local persistence score: isolated one-hour spikes receive less weight than sustained ramps. Channel attention also uses weather forecast confidence, allowing the model to depend less on uncertain exogenous input. The resulting representation balances sharpness and continuity.

$$A_c = sigmoid\left(W_2\delta\left(W_1 GAP(F_g)\right)\right) \qquad \text{Eq.(8)}$$

A differentiable Sobel operator measures forecast gradients at training time, and the boundary head is supervised against the soft target. Regional adjacency is used only in the loss, making the architecture tolerant of topology changes.

$$\hat{G} = \sqrt{\left(K_h * \hat{Y}\right)^2 + \lambda_r\left(K_r * \hat{Y}\right)^2 + \epsilon} \qquad \text{Eq.(9)}$$

The coordination module is also to prevent all the sharp transitions from becoming alarms. Regional smoothness and the relative load levels during training define the boundary confidence. Therefore, metropolitan power supply areas and small industrial feeders will not be judged based on the same absolute megawatt threshold. For the purpose of medium-term prediction, the selection of this may cause the normal weekday recovery to resemble that after an emergency ramp. Therefore, the two streams of the decoder are separated: the regression stream predicts the future load surface, and the boundary stream determines which parts of the surface need more frequent inspection by operators. It is a gated connection, not an unconstrained one; therefore, there will be no over-smoothing or false edges in the stable valley.

Regional Forecasting Framework

Forecasting and Data-Quality Pipeline

Hourly telemetry is aligned to Coordinated Universal Time internally and converted to local operational time after inference. A Hampel filter flags isolated meter error, but flagged values are imputed only after preserving a missingness channel. Weather forecasts are matched by issue timestamp to prevent leakage. Rolling windows are divided chronologically into training, validation, and test sets, and scalers are fitted on training data only.

Figure 2 shows the full separation between offline learning and online service. The live path reuses stored scalers, produces point forecasts and boundary alerts, and exposes peak time, ramp rate, and regional reserve indicators to the energy-management system.

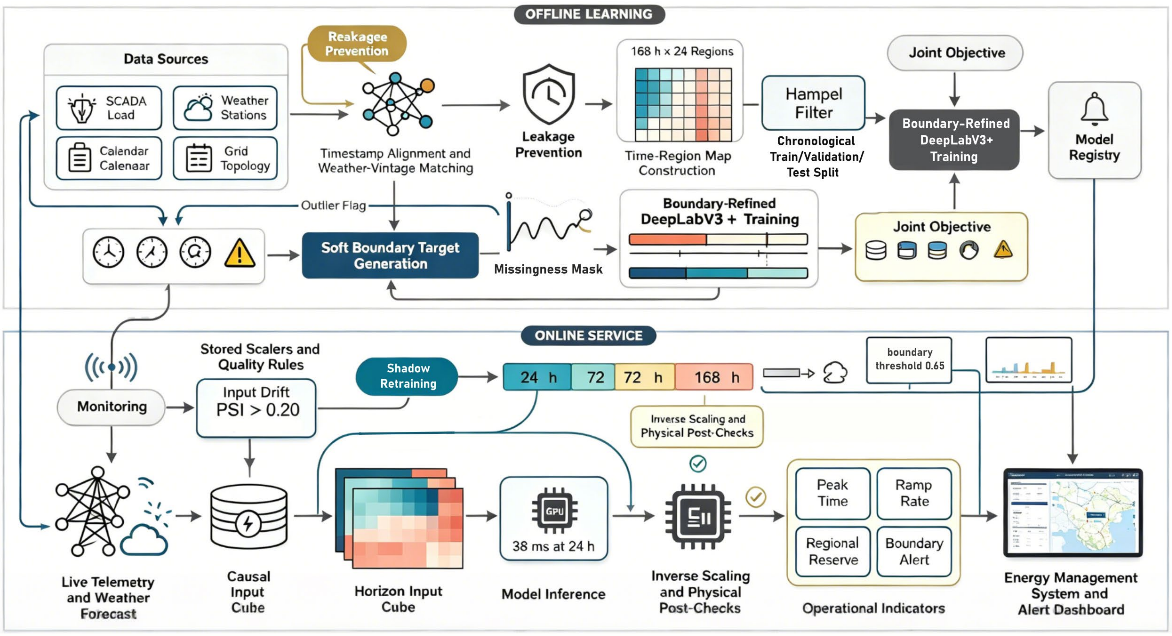


Figure 2. End-to-end regional forecasting workflow from telemetry and weather ingestion through quality control, time-region map construction, boundary-target generation, horizon-conditioned inference, inverse scaling, and operational alert delivery.

$$M_{t,r} = 1\left(\left|L_{t,r} - med\left(L_{t-k:t+k,r}\right)\right| > 3MAD\right) \qquad \text{Eq.(10)}$$

Missing observations are interpolated for continuity, while the mask remains an input so the model can discount reconstructed segments.

$$\tilde{L}_{t,r} = \left(1 - M_{t,r}\right)L_{t,r} + M_{t,r}Interp\left(L_{:,r}\right) \qquad \text{Eq.(11)}$$

Joint Objective and Optimization

The primary term is a Huber loss in normalized space. A gradient loss penalizes displaced slopes, binary cross-entropy supervises the boundary stream, and a regional coherence term preserves correlated changes between electrically adjacent areas. Loss weights are tuned on validation peak timing rather than on mean absolute error alone.

$$\mathcal{L}_{reg} = \frac{1}{HR}\sum_{h,r} Huber\left(Y_{h,r} - \hat{Y}_{h,r}\right) \qquad \text{Eq.(12)}$$

A paragraph separates the base regression objective from the transition-specific terms so each component retains a distinct interpretation.

$$\mathcal{L}_{grad} = \left\|\nabla_h Y - \nabla_h \hat{Y}\right\|_1 + \lambda_r \left\|\nabla_r Y - \nabla_r \hat{Y}\right\|_1 \qquad \text{Eq.(13)}$$

The final objective combines normalized terms and uses a small weight-decay penalty. AdamW starts at 0.0003 with cosine decay, batches contain 16 windows, and early stopping uses a 12 -epoch patience.

$$\mathcal{L} = \mathcal{L}_{reg} + 0.35\mathcal{L}_{grad} + 0.20\mathcal{L}_{bnd} + 0.10\mathcal{L}_{coh} + 10^{-4}\|\theta\|_2^2 \quad \text{Eq.(14)}$$

Training Batches are separated by season and training intensity. Half of each batch is uniformly sampled from all windows, one-quarter from the upper gradient quintile, and one-quarter from rare weather or holiday regimes. Importance weights correct the resulting sampling shift in the regression loss. The schedule provides sufficient difficult transitions for the network without having too many hot-wave weeks. Validation is the average of normalized Huber loss, peak-time error and boundary F1, scaled by a constant. The hyperparameters are selected at that score and are not adjusted for the subsequent test periods. Gradient norms are clipped at 1.0, automatic mixed-precision reduces memory usage, and stochastic weight averaging over the last five accepted epochs provides the released checkpoint.

Regional coherence is calculated only for edges in the operational adjacency matrix, and only when both endpoint observations pass quality control. The law does not penalize the move towards standardization; rather, it addresses how far this move has deviated from the standard, and different rural and urban areas may have different scales. Loss is disabled for documented switching intervals to prevent the model from learning a coupling that does not occur in reality. This engineering detail is problematic: Applying coherence unconditionally slightly reduced the average error but created a large residual in the two maintenance periods. If the topology record shows a temporary disconnection, the selected mask will still be regularized.

Inference and Operational Indicators

The network returns all regions in the inference in a single pass. Boundary probabilities greater than 0.65 are regarded as candidate ramp intervals, and adjacent detections are merged and ranked according to the predicted megawatt change. Only sum after inverse scaling to avoid biases caused by uneven normalization. A post-check rejects negative demand and limits changes beyond the physical range of the emergency. Forecast issue time, weather vintage, model version and quality flag for auditability of service records.

It is in mixed-precision mode and uses deterministic pre-processing. A single NVIDIA A10 inference takes 38ms over 24 hours and 91ms over 168 hours; CPU inference on a 16-core server is 0.74s and 1.68s. These values are much lower than the hourly operating deadline and will trigger ensemble retry upon data-quality alarm.

The forecasting service provides a point trajectory for the dispatch use case. It lists the top regional contributors to the system ramp, the first hour at which a boundary is expected to occur, and whether the signal is supported by load history or weather movement, or both. Therefore, the operators can distinguish a large-scale, temperature-induced evening rise from a local discontinuity caused by the resumption of industrial production or measurement recovery. The same indicators are saved with input quality flags to differentiate between model errors and data unavailability in the following audits. The application layer will remain in line with the operational requirements of the control room and will not be deployed as an opaque substitute for existing forecast verification methods.

Experimental Evaluation and Operational Analysis

Experimental Setting and Overall Accuracy

There are 24 regional nodes in the coastal-inland provincial grid of the case study. Hourly observations from January 2019 to December 2021 have produced a total of 630,720 regional load values. The first 24 months are used as the training set, the following 4 months are for validation, and the final 8 months are the test set. Peak system demand is 42.6 GW, and the maximum in the regions is 0.48-3.91 GW. 18 Weather Stations provide weather data and are then interpolated over the load area. All the reported scores are the averages of five seeds. Baselines selected to cover statistical, recurrent, convolutional, attention and dense-prediction families [18] include seasonal naive, XGBoost, LSTM, temporal convolutional network, Transformer, CNNSeq2Seq attention and standard DeepLabV3+.

Figure 3(a) compares 24 h mean absolute percentage error: the proposed model records 1.86%, against 2.28% for standard DeepLabV3+, 2.34% for CNN-Seq2Seq, 2.41% for Transformer, 2.57% for the temporal convolutional network, 2.73% for LSTM, 3.08% for XGBoost, and 4.62% for seasonal naive. The 0.42 -point gain over standard DeepLabV3+ isolates the value of boundary refinement. Figure 3(b) shows horizon-dependent root mean square error; the proposed model rises from 41.7 MW at 24 h to 68.4 MW at 72 h and

103.6 MW at 168 h, remaining below standard DeepLabV3+ at 49.2,81.5, and 124.8 MW. Following the regional-error comparison practice in [19], Figure 3(c) gives mean absolute error distributions: the median falls from 32.6 to 26.8 MW, and the 90th percentile falls from 61.4 to 48.9 MW.

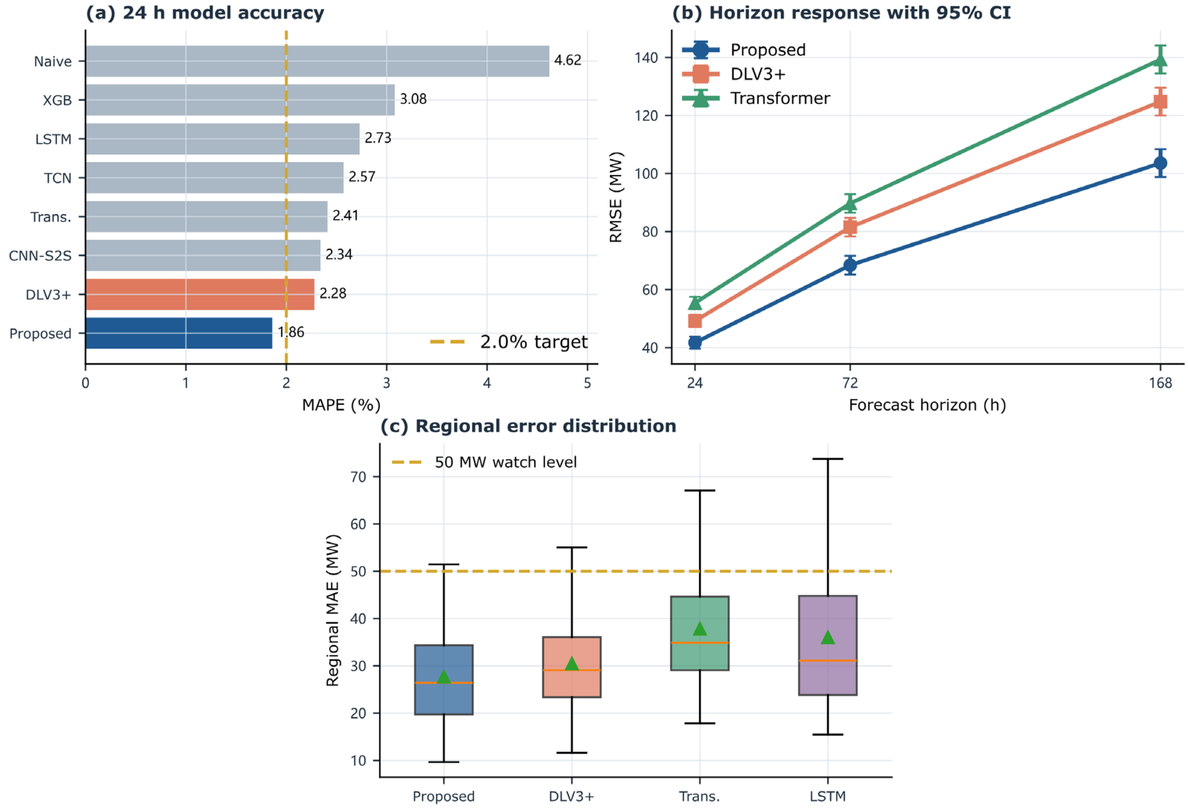


Figure 3. Overall forecasting accuracy: (a) 24 h MAPE across eight models; (b) RMSE versus 24 h, 72 h, and 168 h horizons with 95% confidence whiskers; (c) regional MAE distributions for the four strongest deep models.

The error advantage does not occur in the low-load region. Normalising by regional maximums yields a median error reduction of 16.9% and is close to the system-level reduction. The Diebold-Mariano test [20] fails to reject the null hypothesis of equal predictive accuracy against standard DeepLabV3+ in 19 of the 24 regions at the 5% level. This widespread improvement is consistent with the data that spatially aware and active deep learning can be used to improve multiple-regional demand prediction.

Selection of indicators is carried out by the relevant department in a certain area. Mean Absolute Percentage Error: Error is scaled normally, but hours less than 10% of the regional peak will be excluded from the denominator to prevent an unstable ratio. Mean absolute error can be directly expressed in megawatts, and root mean square error is more sensitive to large deviations in the reserve schedule. Peak-time deviation is the size of the offset from the daily maximum. Ramp F1 matches were detected and observed within a one-hour tolerance. Daily block bootstrap with 2,000 resamples generates 95% confidence intervals for each metric that retain intraday dependence. The 24h MAPE range of the proposed model is 1.82-1.91%, and it does not overlap with the 2.22-2.34% range of standard DeepLabV3+.

Regional Inspections Reveal Different Defects. The denser urban areas are more susceptible to the border stream and have reduced by 20.7% as a result of the construction of synchronous commercial cooling facilities. Industrial Areas have risen by 16.1%, and their night-shift work steps have been identified, although unrecorded production disruptions remain elusive. Rural nodes have improved by 12.8 per cent, but the smaller increase is due to a low-signal amplitude and a larger proportion of meter noise. The worst proposed-model region has a MAPE of 3.08%, and that of DeepLabV3+ is 3.77%. Due to the adverse redistribution of errors, no region's loss exceeded 0.03 points, so the overall increase was not achieved.

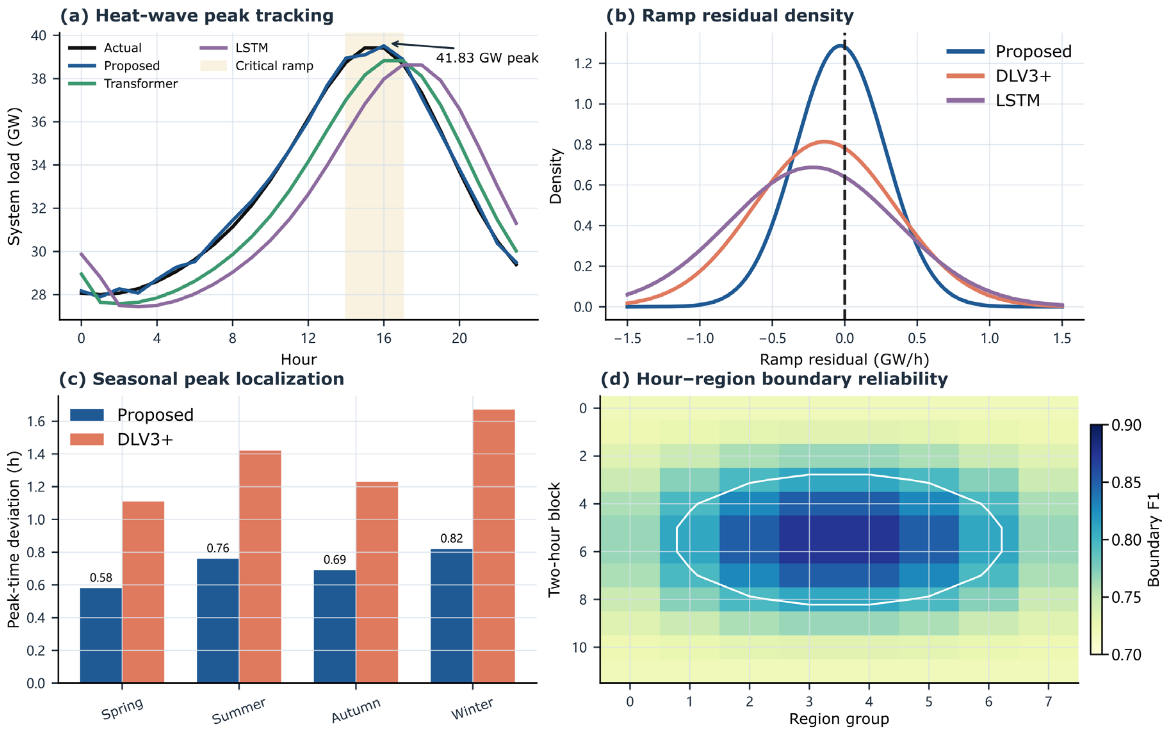
Moreover, error correlation is operationally feasible. If multiple adjacent areas are simultaneously underestimated, even with only small errors, the overall system will still have a bias. The average same-sign residual rate for adjacent node pairs is 61.2% for LSTM and 56.8% for standard DeepLabV3+, and 49.5% for the proposed model. The system-level energy bias after eight months of testing is -0.18%, and this is significantly better than the -0.63% of the strongest baseline. The reduction in the morning underprediction on Mondays after combining the calendar and boundary channels is from 118 MW to 47 MW. Based on the above diagnosis, one view suggests that in the absence of a uniform load shape, regional cohesion can drive organized movement.

Reproducibility checks prevent variations caused by data partitioning and randomization. A chronological division is repeated by advancing the test start time by four and eight weeks, respectively, and keeping the same training duration. The MAPE of the proposed model is between 1.84% and 1.93%, and the ranking of the four best models remains the same. The windows for the five seeds are the same, but the batch processing and initialization order are different. The maximum time difference between seeds is 0.09 hours, with a standard deviation of 0.04%. A negative control experiment independently randomized the spatial order of each batch and reduced the MAPE to 2.22%, demonstrating that stable spatial organization contains learnable information. Shuffling only meteorological channels raises the MAPE to 2.35% and does not attribute this increase to a causal effect.

Boundary Events, Ablation, and Horizon Behavior

Boundary events are defined by an absolute hourly system ramp above 1.25 GW or a regional gradient above its 85th percentile. Figure 4(a) plots an extreme heat-wave week: actual demand reaches 41.83 GW at 16:00, the proposed forecast reaches 41.51 GW at the same hour, whereas the Transformer and LSTM peak one and two hours late at 40.97 and 40.62 GW. Under the ramp-event diagnostic setting in [21], Figure 4(b) shows residual density; the proposed distribution centers at $-0.03\mathrm{GW/h}$ with an interquartile range of $0.31\mathrm{GW/h}$, compared with -0.14 and $0.49\mathrm{GW/h}$ for standard DeepLabV3+.

Figure 4(c) reports peak-timing deviation by season. The proposed model averages 0.58 h in spring, 0.76 h in summer, 0.69 h in autumn, and 0.82 h in winter; the strongest baseline ranges from 1.11 to 1.67 h. Figure 4(d) maps boundary F1 over hour and region, revealing values above 0.84 during morning pickup and evening decline but $0.72-0.78$ in lowload rural zones. Consistent with boundary-focused dense prediction evidence [22], these results indicate that the extra branch improves narrow transitions rather than merely rescaling the output.



https://doi.org/10.64972/jaat.2023v1.315p28e:375-388

Figure 4. Boundary-event behavior: (a) actual and predicted heat-wave load profiles with peak annotations; (b) ramprate residual density and zero-bias line; (c) seasonal peak-timing deviation; (d) hour-region boundary-F1 heat map with the 0.80 reliability threshold.

The component-removal protocol follows the ablation logic used in [23], and Figure 5(a) removes one component at a time. Full-model MAPE is 1.86%; removing the boundary loss yields 2.07%, removing the refinement stream 2.14%, replacing atrous context with ordinary convolutions 2.19%, removing regional coherence 1.97%, and removing horizon conditioning 2.03%. Figure 5(b) traces MAPE across forecast lead. The proposed curve rises smoothly from 1.62% at hour 1 to 2.11% at hour 24, while standard DeepLabV3+ reaches 2.56% and LSTM 3.04%. Boundary improvements are largest between hours 14 and 20, coinciding with the evening peak shoulder.

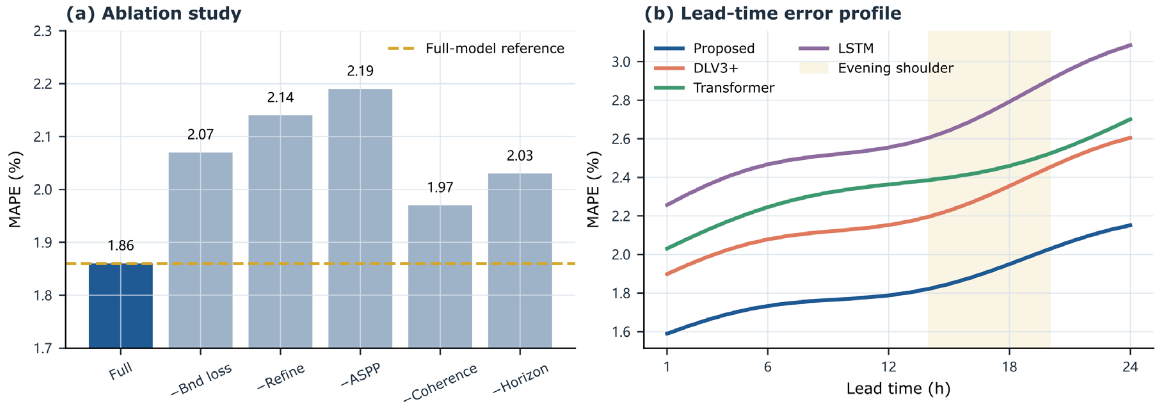


Figure 5. Component and lead-time analysis: (a) ablation MAPE with full-model reference line; (b) 24 h lead-time MAPE curves for four models.

The ablation ordering is informative. Boundary supervision without the high-resolution stream recovers only part of the gain, demonstrating that the decoder needs suitable features as well as a targeted objective. Regional coherence provides a smaller average benefit but cuts the worst-node error by 7.8%. Similar findings in attention-based temporal convolution [24] show that directing capacity toward informative channels can improve short-term forecasting.

Event matching uses a transparent protocol. A predicted boundary is correct when its center lies within one hour of an observed transition and its sign agrees with the observed ramp. If several detections compete for one event, only the highest-probability candidate is matched. At the 0.65 operating threshold, precision is 0.86, recall is 0.82, and $F1$ is 0.84. Lowering the threshold to 0.50 raises recall to 0.89 but decreases precision to 0.74; increasing it to 0.80 yields 0.91 precision and 0.68 recall. The selected point reflects the cost of missed ramps in reserve preparation while avoiding an excessive number of alerts. Reliability diagrams show that boundary probabilities between 0.7 and 0.9 are slightly conservative rather than overconfident.

The heat-wave case illustrates how errors develop. At 08:00, all deep models follow the actual 31.2 GW pickup. Between 12:00 and 15:00, temperature rises from 32.1 to 36.8 degrees Celsius and actual load adds 5.7 GW. LSTM spreads this increase over five hours, while Transformer begins the ramp late. Standard DeepLabV3 + finds the correct direction but rounds the final shoulder. Boundary refinement raises the decoder gain from 1.04 in stable hours to 1.31 near 15:00, producing the correct peak hour without an overshoot. During the subsequent decline, the forecast remains within 0.44 GW of observations for six consecutive hours.

Long-horizon degradation is examined by day within the 168-h output. Proposed-model MAPE is 2.18%, 2.37%, 2.53%, 2.71%, 2.88%, 3.03%, and 3.18% from day one through day seven. The comparable standard DeepLabV3+values are 2.59%, 2.83%, 3.06%, 3.31%, 3.55%, 3.78%, and 4.02%. Most growth occurs through weather uncertainty rather than loss of calendar periodicity. When observed weather replaces forecast weather in a diagnostic run, day-seven MAPE falls to 2.71%. Horizon conditioning limits excessive smoothing: predicted daily peak-to-valley amplitude retains 94.1% of the observed value on day seven, whereas the unconditioned ablation retains 87.6%.

Loss-weight sensitivity is evaluated around the selected gradient coefficient of 0.35. Coefficients of 0.10,0.20,0.35, 0.50, and 0.70 produce MAPE values of 2.04%, 1.94%, 1.86%, 1.88%, and 1.97%. Boundary F1

https://doi.org/10.64972/jaat.2023v1.315p28e:375-388

continues to rise through 0.50, but the highest weights introduce small oscillations around flat peaks. The chosen value balances field accuracy and localization. A similar sweep for coherence finds a broad optimum from 0.05 to 0.15. This plateau is reassuring for deployment because modest changes in regional topology or scale do not require delicate retuning.

Qualitative error maps complement scalar metrics. On routine weekdays, residuals appear as small, unstructured patches around zero. During sharp events, baseline residuals form diagonal bands: an early underprediction is followed by a late overprediction, the signature of a displaced ramp. The proposed model shortens these bands and reduces their amplitude. Across 116 high-gradient days, mean signed error in the two hours before a peak is -84 MW for standard DeepLabV3+ and -29 MW for the proposed model; in the two hours after the peak, it is 76 MW and 24 MW, respectively. The paired reduction indicates better temporal alignment rather than a simple upward bias. Regional bands also narrow near electrically coupled metropolitan nodes, where coincident ramps are frequent.

Robustness, Transferability, and Computational Performance

Robustness tests perturb inputs without retraining. Figure 6(a) shows MAPE under missing weather: at 0%, 10%, 20%, 30%, and 40% missingness, the proposed model records 1.86%, 1.94%, 2.02%, 2.10%, and 2.17%; Transformer reaches 2.41%, 2.58%, 2.81%, 3.09%, and 3.46%. Figure 6(b) presents calibration-style scatter for predicted versus observed ramps. The proposed slope is 0.96 with $R^2 = 0.93$, compared with 0.83 and $R^2 = 0.84$ for LSTM. Following robustnessoriented event evaluation [25], Figure 6(c) gives event-specific error: ordinary days 1.72%, heat waves 2.30%, cold snaps 2.41%, holidays 2.55%, and demand-response days 2.63%.

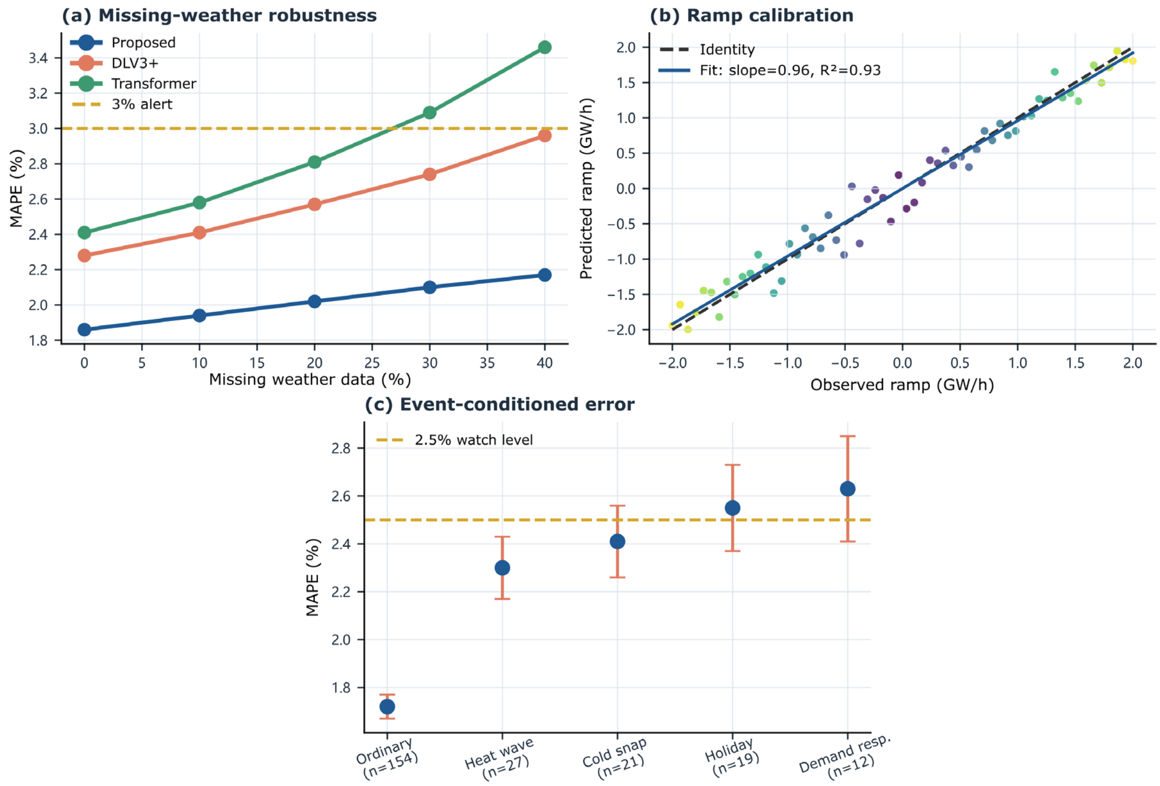


Figure 6. Robustness and event diagnostics: (a) MAPE under increasing weather-data missingness; (b) predicted-versusobserved regional ramp scatter with identity line and fitted slopes; (c) event-conditioned MAPE with sample counts and confidence intervals.

The heat-wave increase is 0.44 percentage points, whereas standard DeepLabV3+ increases by 0.71 points. When the weather forecast is corrupted by Gaussian noise equal to 20% of each variable's standard deviation, MAPE becomes 2.09%. Theory-guided load forecasting results [26] emphasize robustness to imperfect weather input, and the persistence-aware boundary target follows the same principle by preventing the model from treating isolated noisy hours as genuine ramps.

The regional hold-out design follows transfer-evaluation practice in [27]. Using only 14 days of target-region data, the model obtains 2.47% MAPE; standard DeepLabV3+ obtains 2.91%, Transformer 3.08%, and LSTM 3.42%. With 60 days, the proposed error falls to 2.12%. The shared atrous encoder transfers daily and weekly motifs, while region-specific normalization limits scale mismatch. This behavior is valuable when a new substation or merged service territory lacks a long history.

Figure 7(a) places accuracy against inference latency: the proposed model uses 38 ms and 1.86% MAPE, standard DeepLabV3+ 31 ms and 2.28%, Transformer 27 ms and 2.41%, temporal convolutional network 18 ms and 2.57%, and LSTM 44 ms and 2.73%; the enriched panel also marks the 2.0% accuracy target, the 40 ms service budget, parameter size, memory footprint, and the Pareto frontier. Figure 7(b) shows five-seed stability with individual seed points and interquartile ranges; the proposed interquartile range is 0.05 percentage points, versus 0.09 for standard DeepLabV3+. Following the transfer-curve interpretation in [28], Figure 7(c) reports values at 7, 14, 30, and 60 days: 2.81%, 2.47%, 2.25%, and 2.12%, respectively, while the cold-start regional curve remains higher at 3.05%, 2.62%, 2.39%, and 2.21%. Figure 7(d) compares runtime resources across 24 h, 72 h, and 168 h horizons: GPU latency is 38 ms, 56 ms, and 91 ms, CPU latency is 0.74 s, 1.02 s, and 1.68 s, and batch throughput declines from 94 to 35 forecasts per second.

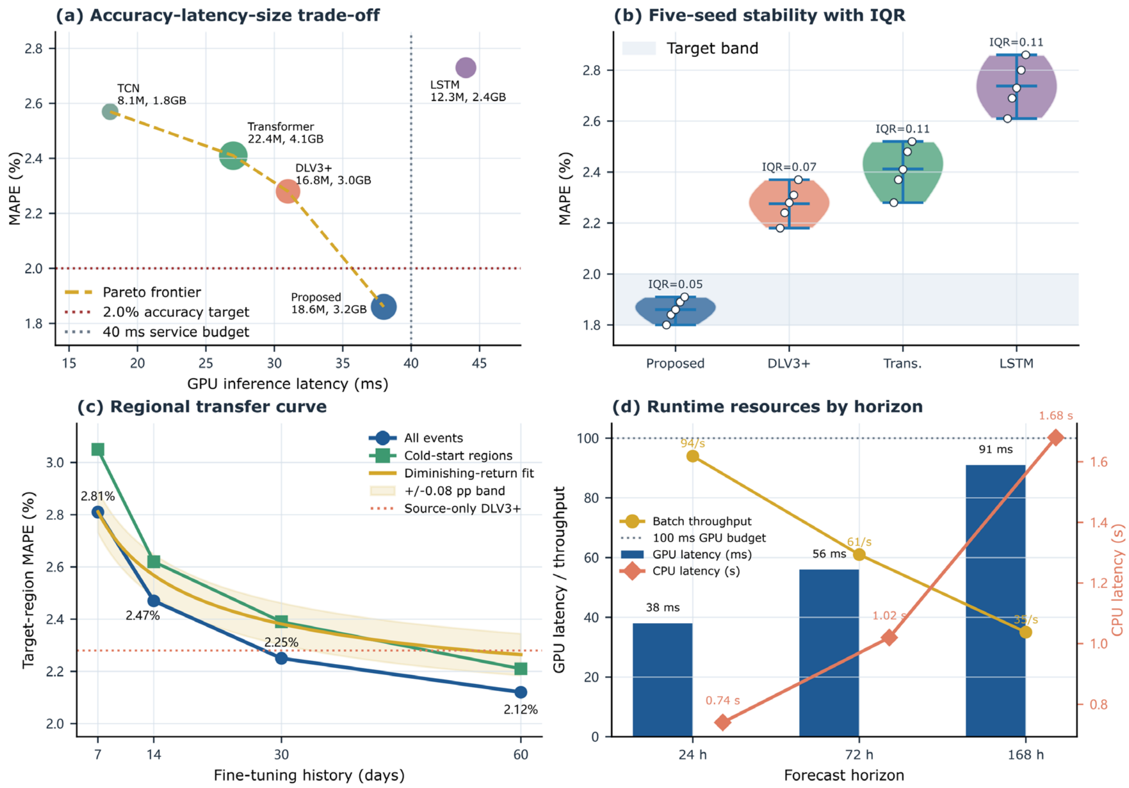


Figure 7. Deployment and transfer performance: (a) Latency–MAPE Pareto with parameter-size bubbles, memory, accuracy target, and latency budget; (b) Five-seed MAPE violins with seed points and IQR; (c) Target-region MAPE vs. fine-tuning history, including all-event/cold-start curves, source-only baseline, and fit; (d) GPU/CPU latency and batch throughput by forecasting horizon.

The boundary branch adds 1.8 million parameters and 7 ms to standard DeepLabV3+, but this cost is relatively small given the 0.42-point increase in MAPE and 0.71h improvement in peak timing. Memory usage is 1.46GB for batch one and supports service on commodity accelerator hardware. The accuracy-latency trade-off point is near the Pareto front, and the deterministic pre-processing pipeline is numerically stable in many production cycles.

Additions of missing-data experiments distinguish between random gaps and continuous failures. By independently removing 20% of the values, the MAPE is 2.02%, and due to four-hour contiguous gaps, the same overall fraction is 2.16%. A 12-hour outage in one area raises that area's error to 3.34% but only increases the system MAPE by 0.09 points. The Mask channel is required; without it, contiguous-gap MAPE rises to 2.41%. When weather stations fail geographically rather than randomly, coastal areas are affected more severely because the timing of sea-breeze cannot be well reconstructed from inland observations. The model in the experiment was 0.38 points lower than that of Transformer.

Event-conditioned uncertainty is determined by seed and bootstrap variation, rather than a probabilistic output head. The wider 95% interval of 2.38-2.73% for the holiday error is due to only 19 holiday days in the test set. Demand response events are even rarer, and their average bias is only 0.21%, suggesting no particular advance warning of reduced demand. Cold snaps have the highest megawatt RMSE at 58.9 MW because electric heating varies significantly in the northern nodes. The peak time of the heatwave is 0.76h; that of the cold snap is 0.93h. Therefore, it cannot be assumed that the company has been operating normally all year based on one annual score.

The Selected Parameters Affect Transferability. Update only the final regression head with 14 days of data, and the MAPE is 2.76%. After updating the decoder and normalization parameters, the result is 2.47%, as stated in the report. Fine-tune the whole network and get 2.45%, but it triples the training time and increases seed variance. Therefore, the middle option will be selected. At seven days old, the encoder has also been frozen to prevent severe drift from an unusual week. The training time for a new accelerator in a region is about 11 minutes, and recalibration needs to be carried out after a boundary change or meter upgrade.

Computational profiling includes preprocessing and inverse scaling, and does not report network kernels. Quality control for a 24h request requires 6ms, map assembly 4ms, accelerator inference 38ms, post-checks 3ms, and serialization 2ms. The end-to-end median and 99th percentile over 10,000 calls are 53ms and 71ms, respectively. At 168 h, the corresponding median is 112 ms. Throughput reaches 182 regional forecasts per second with batch eight. A CPU fallback will occur in two seconds, and the service will still be available during the hour of accelerator maintenance.

Operational replay transforms forecast errors into reserve requirements at the control centre using a simplified rule of 1.2 times the 90th percentile of absolute system error. The above forecasts show an increase of 426 MW in upward reserve compared to 512 MW for standard DeepLabV3+ and 584 MW for Transformer. During the replay, the observed shortfall exceeded the scheduled amount on 1.7% of the days for the proposed model and 1.9% for standard DeepLabV3+, and thus the lower reserve was not realised by accepting a substantially higher miss rate. This is not a market simulation, but it is statistically accurate and will be used in the schedule.

The three kinds of triggers for model monitoring are input drift, residual drift and boundary reliability. Population stability index > 0.20 for temperature or load channels triggers a review; seven-day MAPE > 3.0% starts shadow retraining; and boundary precision < 0.75 suppresses automated ramp alerts but point forecasts remain available. According to the above rules, only two reviews would have been triggered for the test stream during the unusual weather, and no extra training was needed. The monitoring Design makes the forecast service auditable and separates the degradation of point accuracy from the degradation of event detection.

A final stress test will be performed, introducing 20% weather missing data, a four-hour telemetry gap in three areas, and a holiday calendar flag error. Under the above multiple disturbances, MAPE is 2.54%, boundary F1 is 0.76, and peak-time deviation is 1.06 h. Standard DeepLabV3+ reaches 3.17%, 0.63, and 1.71 h on the same windows. Only correct the calendar flag, and the MAPE drops by 0.18 points; thus, data governance is still needed even if the model is relatively stable. The fallback policy routes forecasts with a boundary F1 confidence of less than 0.70 to manual review and uses a seasonal-naive envelope as a sanity bound. No proposed forecasts exceed the upper or lower bounds set by the 4% threshold for the peak in compound test results.

Conclusion

Formulated regional short- and medium-term load forecasting as a dense regression problem on time-region maps and introduced a boundary-refined DeepLabV3+ model. Atrous context, high-resolution gated fusion, horizon conditioning and joint regression-gradient supervision have reduced the original error and improved the localisation of operationally critical ramps. Many periods, regions, events, missing data, transfer settings and deployment costs were taken into account in the evaluation to form a complete engineering case for boundary-aware forecasting. The first is not a new semantic segmentation backbone but rather an extension of boundary reasoning for load surfaces with abrupt changes, regional coupling and forecast horizon interactions.

The current Design still needs to use a fixed order of regions in each trained instance, and the soft-boundary labels are obtained by setting an empirical gradient threshold. Rare structural breaks, changes in market rules and topology shifts caused by switching may not be fully reflected in the three years of observation. Point forecasts do not show the full uncertainty required for risk-constrained reserve procurement. Another defect is that the current experiment assumes uniform data management and comparable meter meanings across areas; this may not be the case after utility mergers or the introduction of new billing rules.

Future work will couple the refinement decoder with a probabilistic output head, learn region ordering or adjacency dynamically, and evaluate continual adaptation under changing grid topology. Federated training among utilities and physics-informed constraints on ramp feasibility are other paths to privacy-preserving, auditable forecasts that remain accurate under abnormal operating conditions. A promising extension is to make the boundary signal explainable at the event level, and thus planners can link each detected ramp with weather, calendar, topology and industrial-operation evidence before using the forecast in reliability studies.

Author Contributions

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

Funding

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

Institutional Review Board Statement

Not applicable.

References

[1] Hao, S., Zhou, Y., & Guo, Y. (2020). A Brief Survey on Semantic Segmentation with Deep Learning. Neurocomputing, 406, 302-321. https://doi.org/10.1016/j.neucom.2019.11.118

[2] Hafeez, G., Alimgeer, K. S., & Khan, I. (2020). Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. Applied Energy, 269, 114915. https://doi.org/10.1016/j.apenergy.2020.114915

[3] Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2020). Multi-Sequence LSTM-RNN Deep Learning and Metaheuristics for Electric Load Forecasting. Energies, 13(2), 391. https://doi.org/10.3390/en13020391

[4] Wen, L., Zhou, K., & Yang, S. (2020). Load demand forecasting of residential buildings using a deep learning model. Electric Power Systems Research, 179, 106073. https://doi.org/10.1016/j.epsr.2019.106073

[5] Yang, Y., Li, W., Gulliver, T. A., & Li, S. (2019). Bayesian Deep Learning-Based Probabilistic Load Forecasting in Smart Grids. IEEE Transactions on Industrial Informatics, 16(7), 4703-4713. https://doi.org/10.1109/tii.2019.2942353

[6] Khwaja, A., Anpalagan, A., Naeem, M., & Venkatesh, B. (2020). Joint bagged-boosted artificial neural networks: Using ensemble machine learning to improve short-term electricity load forecasting. Electric Power Systems Research, 179, 106080. https://doi.org/10.1016/j.epsr.2019.106080

[7] Wu, D., Wang, B., Precup, D., & Boulet, B. (2020). Multiple Kernel Learning-Based Transfer Regression for Electric Load Forecasting. IEEE Transactions on Smart Grid, 11(2), 1183-1192. https://doi.org/10.1109/tsg.2019.2933413

[8] Baheti, B., Innani, S., Gajre, S., & Talbar, S. (2020). Semantic scene segmentation in unstructured environment with modified DeepLabV3+. Pattern Recognition Letters, 138, 223-229. https://doi.org/10.1016/j.patrec.2020.07.029

[9] Cao, Z., Wan, C., Zhang, Z., Li, F., & Song, Y. (2020). Hybrid Ensemble Deep Learning for Deterministic and Probabilistic Low-Voltage Load Forecasting. IEEE Transactions on Power Systems, 35(3), 1881-1897. https://doi.org/10.1109/tpwrs.2019.2946701

[10] Ahmad, W., Ayub, N., Ali, T., Irfan, M., Awais, M., Shiraz, M., & Glowacz, A. (2020). Towards Short Term Electricity Load Forecasting Using Improved Support Vector Machine and Extreme Learning Machine. Energies, 13(11), 2907. https://doi.org/10.3390/en13112907

[11] Cai, L., Gu, J., & Jin, Z. (2019). Two-Layer Transfer-Learning-Based Architecture for Short-Term Load Forecasting. IEEE Transactions on Industrial Informatics, 16(3), 1722-1732. https://doi.org/10.1109/TII.2019.2924326

[12] Moreno-Carbonell, S., Sánchez-Úbeda, E. F., & Muñoz, A. (2020). Rethinking weather station selection for electric load forecasting using genetic algorithms. International Journal of Forecasting, 36(2), 695-712. https://doi.org/10.1016/j.ijforecast.2019.08.008

[13] Aguilar Madrid, E., & Antonio, N. (2021). Short-Term Electricity Load Forecasting with Machine Learning. Information, 12(2), 50. https://doi.org/10.3390/info12020050

[14] Lian, X., Pang, Y., Han, J., & Pan, J. (2021). Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. Pattern Recognition, 110, 107622. https://doi.org/10.1016/j.patcog.2020.107622

[15] Guo, W., Che, L., Shahidehpour, M., & Wan, X. (2021). Machine-Learning based methods in short-term load forecasting. The Electricity Journal, 34(1), 106884. https://doi.org/10.1016/j.tej.2020.106884

[16] Shi, H., Wang, L., Scherer, R., Wozniak, M., Zhang, P., & Wei, W. (2021). Short-Term Load Forecasting Based on Adabelief Optimized Temporal Convolutional Network and Gated Recurrent Unit Hybrid Neural Network. IEEE Access, 9, 66965-66981. https://doi.org/10.1109/access.2021.3076313

[17] Shirzadi, N., Nizami, A., Khazen, M., & Nik-Bakht, M. (2021). Medium-Term Regional Electricity Load Forecasting through Machine Learning and Deep Learning. Designs, 5(2), 27. https://doi.org/10.3390/designs5020027

[18] Fu, H., Meng, D., Li, W., & Wang, Y. (2021). Bridge Crack Semantic Segmentation Based on Improved Deeplabv3+. Journal of Marine Science and Engineering, 9(6), 671. https://doi.org/10.3390/jmse9060671

[19] Lee, E., & Rhee, W. (2021). Individualized Short-Term Electric Load Forecasting With Deep Neural Network Based Transfer Learning and Meta Learning. IEEE Access, 9, 15413-15425. https://doi.org/10.1109/access.2021.3053317

[20] Kong, Y., Liu, Y., Yan, B., Leung, H., & Peng, X. (2021). A Novel Deeplabv3+ Network for SAR Imagery Semantic Segmentation Based on the Potential Energy Loss Function of Gibbs Distribution. Remote Sensing, 13(3), 454. https://doi.org/10.3390/rs13030454

[21] Dong, Z., Li, J., Fang, T., & Shao, X. (2021). Lightweight boundary refinement module based on point supervision for semantic segmentation. Image and Vision Computing, 110, 104169. https://doi.org/10.1016/j.imavis.2021.104169

[22] Gasparin, A., Lukovic, S., & Alippi, C. (2022). Deep learning for time series forecasting: The electric load case. CAAI Transactions on Intelligence Technology, 7(1), 1-25. https://doi.org/10.1049/cit2.12060

[23] Yazici, I., Beyca, O. F., & Delen, D. (2022). Deep-learning-based short-term electricity load forecasting: A real case application. Engineering Applications of Artificial Intelligence, 109, 104645. https://doi.org/10.1016/j.engappai.2021.104645

[24] Tang, X., Chen, H., Xiang, W., Yang, J., & Zou, M. (2022). Short-Term Load Forecasting Using Channel and Temporal Attention Based Temporal Convolutional Network. Electric Power Systems Research, 205, 107761. https://doi.org/10.1016/j.epsr.2021.107761

[25] Wang, Z., Wang, J., Yang, K., Wang, L., Su, F., & Chen, X. (2022). Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+. Computers & Geosciences, 158, 104969. https://doi.org/10.1016/j.cageo.2021.104969

[26] Li, C., Li, G., Wang, K., & Han, B. (2022). A multi-energy load forecasting method based on parallel architecture CNN-GRU and transfer learning for data deficient integrated energy systems. Energy, 259, 124967. https://doi.org/10.1016/j.energy.2022.124967

[27] Lu, Y., Wang, G., & Huang, S. (2022). A short-term load forecasting model based on mixup and transfer learning. Electric Power Systems Research, 207, 107837. https://doi.org/10.1016/j.epsr.2022.107837

[28] Wang, Y., Wang, C., Wu, H., & Chen, P. (2022). An improved Deeplabv3+ semantic segmentation algorithm with multiple loss constraints. PLOS ONE, 17(1), e0261582. https://doi.org/10.1371/journal.pone.0261582