ICI-Time: In-Context Inpainting for Adaptable Time Series Forecasting

Thang Nguyen, Dung Nguyen, Romero Morais, and Truyen Tran

Applied Artificial Intelligence Initiative, Deakin University,
Geelong, Victoria 3216, Australia
minh.t.nguyen@deakin.edu.au

Abstract. We propose ICI-Time, a novel framework that reframes time series forecasting as a visual inpainting task, leveraging the generalisation power of large vision models (LVMs). Unlike methods that require specialised temporal architectures and extensive domain-specific training, ICI-Time transforms time series into structured visual representations (area charts) and applies visual in-context learning, reformulating forecasting as pattern completion within a grid-structured prompt that pre-trained vision transformers can solve without fine-tuning or architectural modification. Temporal dependencies are represented through spatial layout, with a consistent, invertible mapping between numerical and visual domains. Extensive experiments across epidemiology, meteorology, and power systems demonstrate that ICI-Time performs competitively against deep learning baselines and shows promising adaptability under limited-data settings, introducing a new paradigm that bridges temporal and visual domains.

Keywords: Time-series forecasting · Visual prompting · Inpainting · In-context learning

1 Introduction

Time series forecasting is a foundational problem in machine learning, underpinning applications across finance, epidemiology, power systems, and climate modeling. Despite decades of progress, forecasting remains inherently challenging due to complex temporal dependencies, multi-scale variability, non-stationarities, and the scarcity of labeled data in many real-world settings. Deep learning approaches have achieved state-of-the-art results [13, 35], but they typically require domain-specific architectures and extensive training or fine-tuning when adapting to new domains or tasks. This reliance on task-specific engineering limits their scalability and hinders generalisation, especially in low-data regimes.

In contrast, foundation models in language [3, 27] and vision [2] have demonstrated unprecedented generalisation through in-context learning (ICL)—the ability to solve new tasks purely through exposure to examples at inference time, without parameter updates. Recent efforts have explored ICL for temporal data using large language models (LLMs) [36] or time series models [14], but

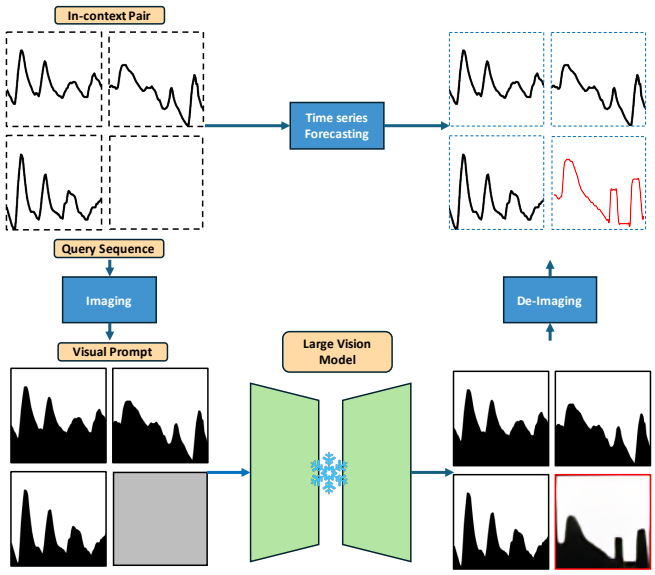

Fig. 1. Visual prompting for time series forecasting via image-based in-context learning. Input sequences are converted into images and arranged in a grid: the first row holds an in-context input–output example pair, the second row holds the query with the missing forecast (gray). A pre-trained large vision model inpaints the missing region, and the prediction (red) is recovered through a de-imaging process.

leveraging large vision models (LVMs) for time series forecasting remains an open challenge, primarily because temporal signals are not natively visual.

In this paper, we introduce ICI-Time (**I**n-**C**ontext **I**npainting for **Time** series), a novel framework that redefines time series forecasting as a promptable visual inpainting task. ICI-Time (i) transforms time series into visual representations (area charts), (ii) assembles input–output example pairs and the query instance into a grid-like visual prompt, (iii) employs a pre-trained LVM to forecast by completing the missing region of the image—analogous to pattern completion, and (iv) translates the generated chart back to numerical time series. As shown in Figure 1, the LVM infers the task from the in-context example pair (x_1, y_1) and inpaints the forecast for the query x_q (shown in red), all without task-specific fine-tuning. Crucially, the LVM is used off-the-shelf, without any modification. This paradigm therefore (i) removes the need to train domain-specific models, (ii) supports rapid adaptation to new forecasting tasks through flexible visual prompting, and (iii) exploits the rich pattern recognition capabilities of LVMs without architectural changes or additional supervision.

Our contributions are fourfold:

– *A cross-modal forecasting framework* that bridges temporal and visual domains, enabling fast adaptation using off-the-shelf vision transformers.

- *A new formulation of visual in-context learning for time series*, extending ICL beyond its NLP roots to a paradigm where temporal prediction is solved through visual prompts and inpainting.
- *A carefully designed bidirectional mapping between time series and visual spaces* that represents temporal dependencies via spatial layout and is invertible, ensuring no loss of forecasting fidelity during transformation.
- *Strong empirical validation* in epidemiology (ILI), meteorology (Weather), and power systems (ETT), where ICI-Time matches strong Transformer-based baselines without any training and is markedly more robust in low-data regimes.

Our findings suggest that visual reasoning models can generalise to temporal tasks when equipped with suitable representations, opening a new research direction in harnessing cross-modal transfer for time series analysis.

2 Related Work

Visual In-context Learning In-context learning (ICL) allows a model to condition at inference time on contextual input–output examples and generate the output for a new input without parameter updates, providing a shortcut to adaptability in AI [5]. While text-based ICL emerged in autoregressive language models, *Visual In-Context Learning* (VICL) trains deep networks to fill in patches of grid-like images [2]. Pioneering works such as Painter [25] and SegGPT [26] showed that generalist vision models can perform diverse tasks—from depth estimation to semantic segmentation—by treating them as inpainting conditioned on visual examples. In the temporal domain, WeatherGFM [33] applied this paradigm to multi-modal weather data using grid-based prompts. However, general-purpose VICL for univariate time series forecasting using simple line plots remains underexplored, and key factors such as informative example selection [32] have been studied for semantic tasks but not for temporal dynamics. Our work bridges this gap by designing a visual prompting scheme that activates the forecasting capability of LVMs without parameter updates.

Foundation Models and Time Series Foundation models have enabled new solutions for time series forecasting [31, 24, 30]. Early work adapted LLMs: Chronos [1] and Time-LLM [8] tokenise time series via quantization or text encoding, but such methods often struggle with the "modality gap" between continuous numerical data and discrete text tokens [16]. A newer wave of "vision-first" models has emerged: VisionTS [4] shows that visual Masked Autoencoders pre-trained on ImageNet can serve as zero-shot forecasters by reconstructing masked time-series images, but requires a pre-training task aligned with forecasting; ViTime [29] trains a foundation model in a binary image metric space for robust probabilistic forecasting. This vision-first trend has since accelerated: VisionTS++ [22] continually pre-trains the visual backbone on large-scale time-series corpora to narrow the modality gap, DMMV [20] fuses decomposition-based multi-modal views with LVMs for long-term forecasting, and OccamVTS [15] and SVTime [21]

distil the predictive priors of LVM forecasters into lightweight networks. These results strengthen the evidence that visual priors transfer to temporal data, yet each still depends on continual pre-training, fine-tuning, or distillation. Unlike these approaches, which require task-aligned pre-training, architectural adaptation, or knowledge transfer into new weights, we investigate the "free lunch" hypothesis [4] using standard line plots, pre-trained generalist models, and in-context visual conditioning—without task-aligned pre-training.

Image-based Representation for Time-series Forecasters Transforming time series into images bypasses the constraints of numerical sequence modeling; a recent survey [16] categorises such transformations into line plots, heatmaps, and spectral images. VisionTS [4] and TimesNet [28] use periodicity-based heatmaps to capture long-term dependencies, but these lose the "shape" of the data, whereas line plots preserve the continuity and topology critical for visual pattern recognition. ViTST [11] converts irregularly sampled series into line graphs for classification, and Time-VLM [34] and VLM-TSC [18] feed line plots to Vision-Language Models, showing that explicit connectivity cues outperform scatter plots or text descriptions; DePlot [12] and ChartLlama [7] further prove that LVMs can extract precise numerical semantics from charts. Most recently, TimeOmni-VL [6] unifies time-series understanding and generation within a single vision–language model and identifies low-loss bidirectional image–series conversion as a prerequisite for numerically faithful generation—independently corroborating a central design principle of our framework—but attains it through large-scale multi-task training on a purpose-built corpus. Yet most existing image-based forecasters [19, 10, 23] train a decoder from scratch or fine-tune the vision backbone, Time-VLM merely augments forecasting with a vision-language model, and unified models such as TimeOmni-VL remain training-intensive. In contrast, we reframe forecasting entirely as visual reasoning: we use a pre-trained visual prompting model [2] as-is—no fine-tuning, no external modules—and design a prompting scheme that leverages its inherent ICL ability to forecast directly from line plot grids.

3 Preliminaries

3.1 Time Series Forecasting

Let $\mathbf{X} \in \mathbb{R}^{C \times L}$ represent a multivariate time series, where L is the total length and C is the number of channels. We partition $\mathbf{X}$ into historical series $\mathbf{X}_I \in \mathbb{R}^{C \times T_I}$ and future series $\mathbf{X}_P \in \mathbb{R}^{C \times T_P}$, with $L = T_I + T_P$. $\mathbf{X}_t^j$ is the series value at the t-th timestep and the j-th channel.

The objective is to develop a predictor $f : \mathbb{R}^{C \times L_I} \to \mathbb{R}^{C \times L_P}$ that maps a lookback window of length L_I to a prediction horizon of length L_P. Here T_I and T_P represent the total available data for training and evaluation, while L_I and L_P define the model's fixed input and output dimensions, where $T_I \gg (L_I + L_P)$.

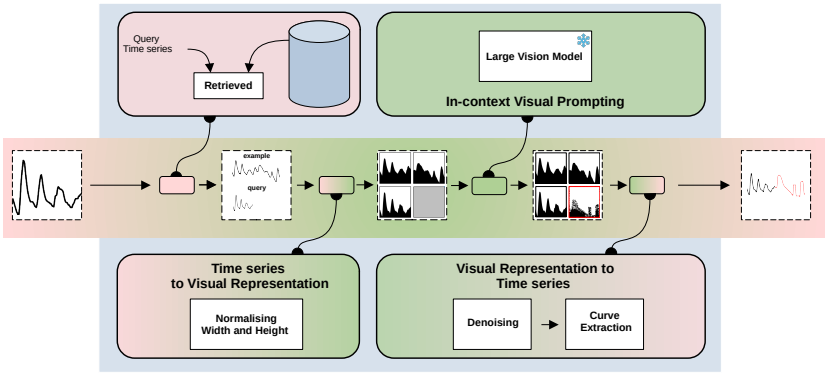

Fig. 2. The ICI-Time framework: (1) example selection from historical data (top-left); (2) time-series-to-image transformation (bottom-left); (3) in-context visual prompting with a pretrained LVM [2] that inpaints the visual prompt (top-right); and (4) image-to-time-series conversion with denoising and curve extraction (bottom-right). The rightmost image shows the original series in black and the forecast in red.

3.2 In-context Learning

In-context Learning (ICL) enables a model to answer a query without task-specific training or fine-tuning: the model is given a sequence of "similar" input–output pairs $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{H}$ before the query $\mathbf{x}_q$, and produces the output $\hat{\mathbf{y}}_q$ directly. This contrasts with standard supervised learning, where input–output pairs are used to train or fine-tune the model.

4 Method

Time series forecasting traditionally requires specialised architectures and training procedures tailored to temporal data. We challenge this paradigm by leveraging the *in-context learning* capabilities of Large Vision Models (LVMs) through visual prompting: time series are transformed into visual representations and the LVM forecasts via inpainting, eliminating the need for domain-specific architectures and model training or tuning, and allowing rapid adaptation to new domains even with limited data. We call this the *in-context inpainting* approach to time series forecasting.

Our framework (Fig. 2) comprises four components: (1) visual case-based selection from historical data, (2) conversion of time series into information-dense images, (3) an off-the-shelf LVM performing in-context inpainting, and (4) a back-conversion mechanism recovering the original time series.

4.1 Visual Case-Based Forecasting

Task examples enable LVMs to learn *in-context*, so it is crucial to select examples relevant to the query. This resembles case-based reasoning—solving new prob-

lems by recalling and adapting similar past problems and their solutions—with the LVM acting as a nonlinear case interpolator in image space.

To enable this, we build a searchable database from $\mathbf{X}_I$ by partitioning the historical series into overlapping windows via a sliding window, each consisting of a lookback component of length L_I and a prediction component of length L_P. For each valid starting position $s \in \{1, \ldots, T_I - (L_I + L_P) + 1\}$:

$$W_s = \{\mathbf{x}_s, \mathbf{y}_s\} = \{\mathbf{X}_I[:, s : s + L_I], \mathbf{X}_I[:, s + L_I : s + L_I + L_P]\}\,. \tag{1}$$

This database $\mathcal{D}_{\text{hist}} = \{W_s\}_{s=1}^{S}$ enables pattern matching during forecasting: given the query window W_q, we compute the Euclidean distance between normalised inputs, $d(W_q, W_s) = ||\bar{\mathbf{x}}_q - \bar{\mathbf{x}}_s||_2$, and select the best example as $W_{s^*} = \arg\min_{W_s} d(W_q, W_s)$.

4.2 Visual Representation of Time Series

The next step transforms numerical time series into visually interpretable formats, using normalization and boundary visualization procedures that balance information preservation and visual clarity.

Image Height The image height parameter normalizes the variety of numerical scales into image sizes typically used in LVMs. Example and query windows are normalised separately.

Example images For the i-th example window $W_i = \{\mathbf{x}_i, \mathbf{y}_i\}$, we apply min-max scaling to both components using shared statistics:

$$\bar{\mathbf{z}} \leftarrow \frac{\mathbf{z} - m_i}{M_i - m_i}, \qquad \mathbf{z} \in \{\mathbf{x}_i, \mathbf{y}_i\}, \tag{2}$$

where $m_i = \min(\mathbf{x}_i^{\min}, \mathbf{y}_i^{\min})$ and $M_i = \max(\mathbf{x}_i^{\max}, \mathbf{y}_i^{\max})$. After normalization, $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{y}}_i$ are plotted into separate figures. The vertical axis extends from $h_{\text{example}}^{\min} = \min(\bar{\mathbf{x}}_i, \bar{\mathbf{y}}_i)$ to $h_{\text{example}}^{\max} = 1.25 \times \max(\bar{\mathbf{x}}_i, \bar{\mathbf{y}}_i)$. The scale $h_{\text{example}} = h_{\text{example}}^{\max} - h_{\text{example}}^{\min}$ serves as explicit prior knowledge transferred to the query.

Query images For a query window $W_q = \{\mathbf{x}_q, \mathbf{y}_q\}$, we normalise using the input's local statistics: $\bar{\mathbf{x}}_q \leftarrow \frac{\mathbf{x}_q - \mathbf{x}_q^{\min}}{\mathbf{x}_q^{\max} - \mathbf{x}_q^{\min}}$ and $\bar{\mathbf{y}}_q \leftarrow \frac{\mathbf{y}_q - \mathbf{x}_q^{\min}}{\mathbf{x}_q^{\max} - \mathbf{x}_q^{\min}}$ for the unknown target. The vertical axis range is $h_{\text{query}}^{\min} = \min(\bar{\mathbf{x}}_q)$ to $h_{\text{query}}^{\max} = h_{\text{query}}^{\min} + h_{\text{example}}$, which ensures visual consistency.

Image Width Our implementation uses fixed 224×224 pixel images. To handle varying sequence lengths, we adjust the x-axis limits: $\text{xlim}_{in} = [1, L_I]$ for lookback components and $\text{xlim}_{out} = [L_I + 1, L_I + L_P]$ for prediction components.

The invertibility of our visual representation function $g : \mathbb{R}^{C \times L_P} \to \mathcal{I}$ necessitates preserving the parameters $\theta_i = \{\mathbf{x}_i^{\min}, \mathbf{x}_i^{\max}\}$ so that the inverse $g^{-1} : \mathcal{I} \to \mathbb{R}^{C \times L_P}$ can denormalize visual predictions via $\hat{\mathbf{y}}_i = \hat{\mathbf{y}}_i^{\text{norm}} \cdot (\mathbf{x}_i^{\max} - \mathbf{x}_i^{\min}) + \mathbf{x}_i^{\min}$.

Table 1. Summary Statistics of Benchmark Datasets.

Characteristic	ILI	Weather	ETTh1/2	ETTm1/2
Number of Features	7	21	7	7
Number of Timesteps	966	52,696	17,420	69,680

4.3 Visual Prompting and Time Series Recovery

The example and query images constitute a visual prompt; the LVM fills the empty region by producing an image $\hat{\mathbf{I}}$ via in-context learning, capturing forecasting patterns from the examples. We then recover the numerical series $\widehat{\mathbf{y}}_q \in \mathbb{R}^{C \times L_P}$ from $\hat{\mathbf{I}}$ in two steps.

Image Denoising We binarize, $\hat{\mathbf{I}}_{\text{binary}}(r, c) = 1$ if $\hat{\mathbf{I}}(r, c) > \tau$, apply a bitwise NOT to obtain $\hat{\mathbf{I}}_{\text{inv}}$, then apply morphological opening, $\hat{\mathbf{I}}_{\text{opened}} = (\hat{\mathbf{I}}_{\text{inv}} \ominus \mathbf{K}) \oplus \mathbf{K}$, where $\mathbf{K}$ is a 3×3 structuring element. The final clean image is $\hat{\mathbf{I}}_{\text{clean}} = \text{NOT}(\hat{\mathbf{I}}_{\text{opened}})$.

Boundary Curve Extraction We extract $\widehat{\mathbf{y}}_q$ by finding the uppermost foreground pixel in each column: for a cleaned image of height h, the boundary coordinate is $y_i = h - \min\{r \mid \hat{\mathbf{I}}_{\text{clean}}(r, c_i) = 1\} - 1$. To ensure continuity, we interpolate $f(c)$ such that $f(c_i) = y_i$ and sample the curve at L_P equidistant points to obtain the final sequence $\widehat{\mathbf{y}}_q$.

5 Experimental Results

5.1 Datasets

We evaluate ICI-Time across three diverse and widely benchmarked domains (summarised in Table 1): *ILI*[1] (Influenza-Like Illness), weekly influenza-like illness patient data collected by the US CDC from 2002 to 2021, whose clear seasonal patterns and long historical record make it particularly valuable for evaluating retrieval-based strategies; *Weather*[2], 21 meteorological indicators such as air temperature and humidity, recorded at 10-minute intervals throughout 2020; and *ETT*[3] (Electricity Transformer Temperature), series from two electric transformers at 15-minute ('m') and hourly ('h') resolutions, yielding four datasets: ETTh1, ETTh2, ETTm1, and ETTm2.

5.2 Experimental Settings

We adopt the data split setting from Nie et al. [17] with one key modification: we merge the training and validation data into a single database as we do not

[1] https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html
[2] https://www.bgc-jena.mpg.de/wetter/
[3] https://github.com/zhouhaoyi/ETDataset

need to train or fine-tune a model, thereby enlarging our dataset. We use a look-back window $L = 96$ for our model and all Transformer-based baselines. Prediction lengths follow Nie et al. [17], with $T \in \{24, 36, 48, 60\}$ for ILI and $T \in \{96, 192, 336, 720\}$ for the other datasets. We adopt channel independence, forecasting each variable separately, a technique proven effective in deep learning approaches [17, 4, 29].

We report Mean Squared Error (MSE) and Mean Absolute Error (MAE), comparing against Transformer-based baselines: Informer [35] (ProbSparse self-attention), Pyraformer [13] (pyramid attention), and LogTrans [9] (log-sparse attention). Baseline results are sourced from Nie et al. [17] when available, with additional experiments conducted to fill gaps. All experiments maintain consistent configurations to ensure fair comparison.

5.3 Forecasting Performance

Sanity check: We implemented a nearest-neighbour baseline that simply reuses the first example from the input sequence as the prediction for all future time steps, to test whether LVMs simply copy the example over. Its results are poor compared to Transformer-based models and to ICI-Time, confirming that the generalisation power of LVMs comes from leveraging their vast source of visual patterns.

Full-data Results Forecasting results are presented in Table 2. ICI-Time outperforms the baselines in most cases, especially on the MAE metric (23 out of 24 cases, with the remaining case being second best). This demonstrates that off-the-shelf visual in-context models can perform competitive time-series forecasting without training. The MSE metric is sensitive to noise and more reflective of the training square-loss function; in our case, noise can be suppressed by averaging over multiple runs, each using a different near-optimal prompting example.

5.4 Analysis of Design Choices

We ablate the key design decisions in ICI-Time: the height and width settings of the visual encoding, and the post-processing applied during time series recovery. Table 3 reports results averaged over all prediction horizons; each ablation column replaces exactly one component of the full model.

Height Settings The vertical axis encodes value magnitude. We compare our proposed h_{transfer} (Section 4.2), which transfers the height scale from examples to queries via $h^{\max}_{\text{query}} = h^{\min}_{\text{query}} + h_{\text{example}}$, against $h_{1.5}$, a fixed scaling factor of 1.5 for all inputs. h_{transfer} consistently outperforms $h_{1.5}$ across ETT and Weather, reducing the average MSE by 9.2% to 18.5% (largest on ETTh1), suggesting that maintaining visual scale consistency between examples and queries enables more effective pattern recognition. ILI is the exception, where the fixed scaling attains a lower average error.

Table 2. *Per-horizon* performance comparison of various forecasting methods on full training data. Best results are in **bold**. The smaller the better.

Dataset	Prediction	ICI-Time	Informer	Pyraformer	LogTrans	ICI-Time	Informer	Pyraformer	LogTrans
		MSE				MAE			
ETTh1	96	0.812	0.941	**0.774**	0.878	**0.598**	0.769	0.672	0.740
	192	1.116	1.007	**0.797**	1.037	0.718	0.786	**0.680**	0.824
	336	1.228	**1.038**	1.205	1.238	**0.746**	0.784	0.897	0.932
	720	1.291	1.144	**0.977**	1.135	**0.757**	0.857	0.788	0.852
ETTh2	96	**0.288**	1.549	0.645	2.116	**0.357**	0.952	0.597	1.197
	192	**0.402**	3.792	0.788	4.315	**0.426**	1.542	0.683	1.635
	336	**0.449**	4.215	0.907	1.124	**0.455**	1.642	0.747	1.604
	720	**0.419**	3.656	0.963	3.188	**0.466**	1.619	0.783	1.540
ETTm1	96	0.578	0.626	**0.543**	0.600	**0.471**	0.560	0.510	0.546
	192	0.673	0.725	**0.557**	0.837	**0.527**	0.619	0.537	0.700
	336	**0.749**	1.005	0.754	1.124	**0.570**	0.741	0.655	0.832
	720	1.059	1.133	**0.908**	1.153	**0.687**	0.845	0.724	0.820
ETTm2	96	**0.198**	0.355	0.435	0.768	**0.288**	0.462	0.507	0.642
	192	**0.241**	0.595	0.730	0.989	**0.318**	0.586	0.673	0.757
	336	**0.324**	1.270	1.201	1.334	**0.372**	0.871	0.845	0.872
	720	**0.379**	3.001	3.625	3.048	**0.412**	1.267	1.451	1.328
Weather	96	**0.219**	0.354	0.896	0.458	**0.232**	0.405	0.556	0.490
	192	**0.280**	0.419	0.622	0.658	**0.273**	0.434	0.624	0.589
	336	**0.399**	0.583	0.739	0.797	**0.339**	0.543	0.753	0.652
	720	**0.443**	0.916	1.004	0.869	**0.410**	0.705	0.934	0.675
ILI	24	3.136	4.657	**1.420**	4.480	**1.068**	1.449	2.012	1.444
	36	**2.875**	4.650	7.394	4.799	**1.014**	1.463	2.031	1.467
	48	**3.572**	5.004	7.551	4.800	**1.098**	1.542	2.057	1.468
	60	**2.709**	5.071	7.662	5.278	**0.985**	1.543	2.100	1.560

Width Settings The horizontal axis represents time. Our proposed w_{diff} uses distinct temporal resolutions for input and target images ($L_I/224$ and $L_P/224$ respectively), preserving the native resolution of each component, whereas w_{even} uses $L_P/224$ uniformly for both, shifting the input to the rightmost position. w_{diff} outperforms w_{even} on all six datasets, reducing the average MSE by 1.8% to 5.1% on ETT and Weather and by 10.7% on ILI, indicating that preserving the native temporal resolution of the input consistently benefits forecasting.

Post-processing A critical challenge in recovering time series from generated images is the *disconnection problem*—a large forecasting error at the first prediction step, at the boundary between the input and the predicted values. We apply a Gaussian smoothing decay, weighting $w_i = \exp(-0.5 \cdot (i/\sigma)^2)$ with $\sigma = \text{window}/3$, which smoothly blends the last input point into the predictions over approximately 10 time steps. Compared with raw recovery, smoothing improves the average MSE on all six datasets, by 0.7% to 2.2%, with the largest gain on ILI.

Few-shot adaptation We evaluate few-shot adaptation by restricting all methods to the first 1%, 5%, or 10% of the training data, following Zhou et al. [36]; this simulates forecasting well beyond the temporal range of the training data. Informer and Pyraformer are trained under the same restricted protocol. Since ICI-Time leverages historical samples directly as in-context examples, we imple-

Table 3. Ablation of design choices, *averaged* over all prediction horizons. Each ablation column replaces one component of the full model (ICI-Time = h_{transfer} + w_{diff} + smoothing). Best results are in **bold**. The smaller the better.

Dataset	$h_{1.5}$	w_{even}	Raw	ICI-Time	$h_{1.5}$	w_{even}	Raw	ICI-Time
	MSE				MAE			
ETTh1	1.365	1.151	1.126	**1.112**	0.780	0.719	0.712	**0.705**
ETTh2	0.446	0.411	0.394	**0.390**	0.462	0.443	0.430	**0.426**
ETTm1	0.860	0.795	0.777	**0.765**	0.597	0.573	0.570	**0.564**
ETTm2	0.316	0.293	0.288	**0.286**	0.368	0.354	0.351	**0.348**
Weather	0.369	0.341	0.340	**0.335**	0.324	0.315	0.316	**0.314**
ILI	**2.934**	3.440	3.141	3.073	**0.996**	1.127	1.064	1.041

mented a masking procedure to prevent data leakage: when retrieving examples at test time, the initial portion of the test inputs $\mathbf{X}_I$ that would be unavailable in a real deployment is replaced with zeros, so distances are computed against zero-filled rather than actual historical values. This preserves the benefit of retrieval while maintaining the integrity of the few-shot conditions.

The results are presented in Table 4 (10%), Table 5 (5%), and Table 6 (1%), all reporting *per-horizon* results; at 5% and 1%, only the horizons providing sufficient data to train the baselines are included. ICI-Time maintains robust performance across restricted data regimes, often achieving errors that are multiples lower than the baselines. For instance, on ETTh2 (96 pred. length) with 10% data, ICI-Time achieves an MSE of 0.317, while Informer and Pyraformer struggle at 4.047 and 4.065, respectively. At 1%, only ETTm1, ETTm2, and Weather provide sufficient sequence length to train the Transformer-based baselines; even in these extreme cases ICI-Time dominates, e.g., on ETTm2 (96 pred. length) it maintains an MSE of 0.208, whereas Informer's error increases to 1.984.

Figure 3 plots the MAE as a function of training data size for the Weather dataset (96 pred. length). As data decreases from 10% to 1%, the MAE of Informer increases from 0.389 to 0.514 (+32.1%) and Pyraformer from 0.360 to 0.487 (+35.3%), whereas ICI-Time remains remarkably stable, moving only from 0.234 to 0.242 (+3.4%). This highlights the superior data efficiency and adaptability of ICI-Time in low-resource settings.

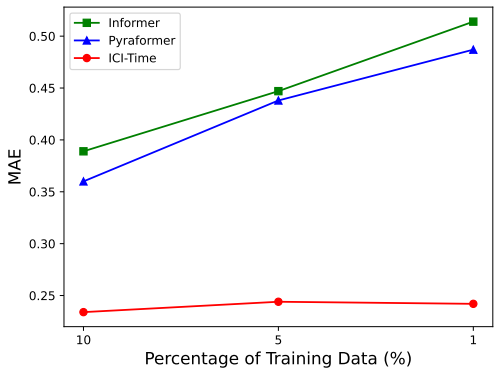

Fig. 3. MAE of ICI-Time, Informer, and Pyraformer on the Weather dataset across percentages of training data, at prediction length $T = 96$.

Table 4. *Per-horizon* few-shot performance of ICI-Time, Informer, and Pyraformer using only the first 10% of training data. Best results are in **bold**. The smaller the better.

Dataset	Prediction	ICI-Time	Informer	Pyraformer	ICI-Time	Informer	Pyraformer
		MSE			MAE		
ETTh1	96	**1.046**	2.113	1.998	**0.687**	1.170	0.984
	192	**1.310**	1.913	1.987	**0.793**	1.000	0.978
	336	**1.407**	2.099	2.382	**0.834**	1.021	1.053
	720	**1.281**	2.332	1.550	**0.800**	1.067	0.909
ETTh2	96	**0.317**	4.047	4.065	**0.376**	1.602	1.591
	192	**0.447**	3.996	4.452	**0.445**	1.533	1.661
	336	**0.511**	4.082	4.797	**0.488**	1.568	1.721
	720	**0.461**	4.867	4.404	**0.480**	1.736	1.650
ETTm1	96	**0.743**	1.641	1.570	**0.560**	0.937	0.934
	192	**0.888**	2.037	1.726	**0.622**	1.104	0.996
	336	**0.971**	2.188	1.998	**0.655**	1.126	1.017
	720	**1.140**	2.799	1.920	**0.727**	1.273	1.029
ETTm2	96	**0.195**	3.080	2.142	**0.291**	1.376	1.164
	192	**0.222**	3.440	3.031	**0.310**	1.449	1.393
	336	**0.323**	2.944	2.479	**0.374**	1.349	1.258
	720	**0.384**	4.052	3.064	**0.416**	1.599	1.377
Weather	96	**0.232**	0.357	0.288	**0.234**	0.389	0.360
	192	**0.292**	0.425	0.364	**0.277**	0.427	0.400
	336	**0.398**	0.720	0.465	**0.339**	0.534	0.439
	720	0.478	0.698	**0.466**	**0.411**	0.528	0.436
ILI	24	**4.412**	7.675	8.341	**1.268**	2.019	2.142
	36	**4.491**	7.595	7.606	**1.308**	2.020	2.018
	48	**4.757**	7.665	7.591	**1.320**	2.037	2.019
	60	**3.969**	8.118	7.950	**1.261**	2.114	2.080

Table 5. *Per-horizon* few-shot performance of ICI-Time, Informer, and Pyraformer using only the first 5% of training data; only the prediction horizons providing sufficient data to train the Transformer-based baselines are shown. Best results are in **bold**. The smaller the better.

Dataset	Prediction	ICI-Time	Informer	Pyraformer	ICI-Time	Informer	Pyraformer
		MSE			MAE		
ETTh1	96	**1.176**	1.802	2.225	**0.743**	0.976	0.998
	192	**1.416**	1.770	1.869	**0.853**	0.972	0.957
	336	**1.353**	2.069	1.478	**0.833**	1.015	0.907
ETTh2	96	**0.316**	3.720	4.609	**0.375**	1.546	1.706
	192	**0.444**	3.884	4.946	**0.448**	1.532	1.758
	336	**0.501**	4.319	3.898	**0.482**	1.624	1.549
ETTm1	96	**1.140**	1.690	1.669	**0.667**	0.996	0.941
	192	**1.404**	1.828	1.882	**0.749**	1.033	1.004
	336	**1.402**	2.115	2.074	**0.765**	1.087	1.003
	720	**1.177**	2.180	2.062	**0.740**	1.084	1.065
ETTm2	96	**0.209**	2.707	2.197	**0.304**	1.304	1.185
	192	**0.232**	3.084	2.567	**0.323**	1.398	1.285
	336	**0.337**	2.917	2.548	**0.382**	1.365	1.268
	720	**0.382**	3.573	3.368	**0.417**	1.502	1.442
Weather	96	**0.247**	0.441	0.405	**0.244**	0.447	0.438
	192	**0.308**	0.563	0.441	**0.285**	0.500	0.448
	336	**0.409**	0.646	0.444	**0.352**	0.540	0.430
	720	0.486	0.578	**0.471**	**0.415**	0.519	0.451
ILI	24	**4.621**	7.453	7.707	**1.312**	2.002	2.046

Table 6. *Per-horizon* few-shot performance using only the first 1% of training data; only the datasets shown provide sufficient data to train the Transformer-based baselines. Best results are in **bold**. The smaller the better.

Dataset	Prediction	ICI-Time	Informer	Pyraformer	ICI-Time	Informer	Pyraformer
		MSE			MAE		
ETTm1	96	**0.960**	1.682	1.851	**0.641**	0.963	1.045
	192	**1.296**	1.743	1.526	**0.743**	1.004	0.922
	336	**1.381**	1.405	1.386	**0.789**	0.895	0.894
ETTm2	96	**0.208**	1.984	2.444	**0.308**	1.118	1.264
	192	**0.234**	3.631	3.029	**0.324**	1.478	1.402
	336	**0.333**	3.340	3.171	**0.380**	1.446	1.404
Weather	96	**0.225**	0.511	0.499	**0.242**	0.514	0.487
	192	**0.266**	0.690	0.587	**0.280**	0.592	0.537
	336	**0.366**	0.729	0.871	**0.339**	0.645	0.724

6 Conclusion

We have shown that visual reasoning can be successfully adapted to model complex temporal dynamics. By transforming time series into structured images and applying in-context inpainting with pre-trained vision models, ICI-Time bypasses the need for specialised temporal architectures and costly training or fine-tuning, while achieving competitive forecasting accuracy across diverse domains. Beyond accuracy, our results reveal a broader insight: pre-trained visual models, combined with carefully designed representations, can generalise far beyond their original modalities, challenging conventional boundaries between temporal and visual modelling. Future research may explore more advanced retrieval strategies for in-context examples, and extensions to time series anomaly detection (e.g., using forecasting error for anomaly scoring) and classification (e.g., encoding classes as visual objects).

Data Availability. The datasets used in this study are publicly available on the internet at `https://github.com/thuml/Autoformer`.

References

1. Ansari, A.F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S.S., Pineda Arango, S., Kapoor, S., et al.: Chronos: Learning the language of time series. Trans. Mach. Learn. Res. (2024)
2. Bar, A., Gandelsman, Y., Darrell, T., Globerson, A., Efros, A.: Visual prompting via image inpainting. Advances in Neural Information Processing Systems **35**, 25005–25017 (2022)
3. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. Advances in neural information processing systems **33**, 1877–1901 (2020)
4. Chen, M., Shen, L., Li, Z., Wang, X.J., Sun, J., Liu, C.: VisionTS: Visual masked autoencoders are free-lunch zero-shot time series forecasters. In: Forty-second International Conference on Machine Learning (2025), `https://openreview.net/forum?id=5DSj3MfWrB`
5. Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Chang, B., et al.: A survey on in-context learning. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. pp. 1107–1128 (2024)
6. Guan, T., Pan, S., Barthelemy, J., Li, Z., Cai, Y., Alippi, C., Jin, M., Pan, S.: Timeomni-vl: Unified models for time series understanding and generation. In: Proceedings of the 43rd International Conference on Machine Learning (ICML) (2026)
7. Han, Y., Zhang, C., Chen, X., Yang, X., Wang, Z., Yu, G., Fu, B., Zhang, H.: Chartllama: A multimodal llm for chart understanding and generation. In: arXiv preprint arXiv:2311.16483 (2023)
8. Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J., Shi, X., Chen, P.Y., Liang, Y., Li, Y.f., Pan, S., et al.: Time-llm: Time series forecasting by reprogramming large language models. In: Proceedings of the 12th International Conference on Learning Representations (ICLR) (2024)

9. Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.X., Yan, X.: Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 32 (2019)
10. Li, Z., Li, S., Yan, X.: Time series as images: Vision transformer for irregularly sampled time series. Advances in Neural Information Processing Systems **36**, 49187–49204 (2023)
11. Li, Z., Li, S., Yan, X.: Time series as images: Vision transformer for irregularly sampled time series. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 36, pp. 49187–49204 (2023)
12. Liu, F., Julian, J., Eisenschlos, J.M., Krichene, Walid andin, F., Collier, N.: Deplot: One-shot visual language reasoning by plot-to-table translation. In: Findings of the Association for Computational Linguistics: ACL 2023. pp. 10381–10399 (2023)
13. Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A.X., Dustdar, S.: Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In: Proceedings of the 10th International Conference on Learning Representations (ICLR) (2022)
14. Lu, J., Sun, Y., Yang, S.: In-context time series predictor. In: Proceedings of the 13th International Conference on Learning Representations (ICLR) (2025)
15. Lyu, S., Zhong, S., Ruan, W., Liu, Q., Wen, Q., Xiong, H., Liang, Y.: Occamvts: Distilling vision models to 1% parameters for time series forecasting. arXiv preprint arXiv:2508.01727 (2025)
16. Ni, J., Zhao, Z., Shen, C., Tong, H., Song, D., Cheng, W., Luo, D., Chen, H.: Harnessing vision models for time series analysis: A survey. In: Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI). pp. 10612–10620 (2025). https://doi.org/10.24963/ijcai.2025/1178
17. Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. In: Proceedings of the 11th International Conference on Learning Representations (ICLR) (2023)
18. Prithyani, V., Mohammed, M., Gadgil, R., Buitrago, R., Jain, V., Chadha, A.: On the feasibility of vision-language models for time-series classification. In: Proceedings of the 59th Hawaii International Conference on System Sciences (HICSS). pp. 1422–1431 (2026)
19. Semenoglou, A.A., Spiliotis, E., Assimakopoulos, V.: Image-based time series forecasting: A deep convolutional neural network approach. Neural Networks **157**, 39–53 (2023). https://doi.org/10.1016/j.neunet.2022.10.006
20. Shen, C., Yu, W., Zhao, Z., Song, D., Cheng, W., Chen, H., Ni, J.: Multi-modal view enhanced large vision models for long-term time series forecasting. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 38 (2025)
21. Shen, C., Zhao, Z., Tong, H., Song, D., Luo, D., Wen, Q., Ni, J.: Svtime: Small time series forecasting models informed by "physics" of large vision model forecasters. arXiv preprint arXiv:2510.09780 (2025)
22. Shen, L., Chen, M., Liu, X., Fu, H., Ren, X., Sun, J., Li, Z., Liu, C.: Visionts++: Cross-modal time series foundation model with continual pre-trained vision backbones. arXiv preprint arXiv:2508.04379 (2025)
23. Sood, S., Zeng, Z., Cohen, N., Balch, T., Veloso, M.: Visual time series forecasting: an image-driven approach. In: Proceedings of the Second ACM International Conference on AI in Finance. pp. 1–9 (2021)
24. Su, J., Jiang, C., Jin, X., Qiao, Y., Xiao, T., Ma, H., Wei, R., Jing, Z., Xu, J., Lin, J.: Large language models for forecasting and anomaly detection: A systematic literature review. CoRR **abs/2402.10350** (2024).

https://doi.org/10.48550/ARXIV.2402.10350, https://doi.org/10.48550/arXiv.2402.10350
25. Wang, X., Zhang, W., Cao, Y., Shen, C., Huang, T.: Images speak in images: A generalist painter for in-context visual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6830–6839 (2023)
26. Wang, X., Zhang, X., Cao, Y., Wang, W., Shen, C., Huang, T.: Seggpt: Segmenting everything in context. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1130–1140 (2023)
27. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al.: Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems **35**, 24824–24837 (2022)
28. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis. In: International Conference on Learning Representations (ICLR) (2023)
29. Yang, L., Wang, Y., Fan, X., Cohen, I., Chen, J., Zhang, Z.: Vitime: Foundation model for time series forecasting powered by vision intelligence. Transactions on Machine Learning Research (2025)
30. Ye, J., Zhang, W., Yi, K., Yu, Y., Li, Z., Li, J., Tsung, F.: A survey of time series foundation models: Generalizing time series representation with large language model. arXiv preprint arXiv:2405.02358 (2024)
31. Zhang, X., Chowdhury, R.R., Gupta, R.K., Shang, J.: Large language models for time series: a survey. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. IJCAI '24 (2024). https://doi.org/10.24963/ijcai.2024/921
32. Zhang, Y., Zhou, K., Liu, Z.: What makes good examples for visual in-context learning? Advances in Neural Information Processing Systems **36**, 17773–17794 (2023)
33. Zhao, X., Zhou, Z., Zhang, W., Liu, Y., Chen, X., Gong, J., Chen, H., Fei, B., Chen, S., Ouyang, W., et al.: Weathergfm: Learning a weather generalist foundation model via in-context learning. In: Proceedings of the 13th International Conference on Learning Representations (ICLR) (2025)
34. Zhong, S., Ruan, W., Jin, M., Li, H., Wen, Q., Liang, Y.: Time-vlm: Exploring multimodal vision-language models for augmented time series forecasting. In: Proceedings of the 42nd International Conference on Machine Learning (ICML) (2025)
35. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI). vol. 35, pp. 11106–11115 (2021). https://doi.org/10.1609/aaai.v35i12.17325
36. Zhou, T., Niu, P., Wang, X., Sun, L., Jin, R.: One fits all: Power general time series analysis by pretrained lm. In: Advances in Neural Information Processing Systems (NeurIPS) (2023)