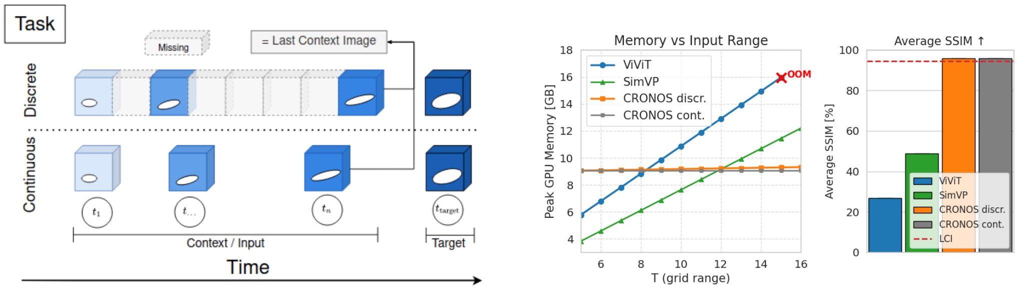
CRONOS: CONTINUOUS TIME RECONSTRUCTION FOR 4D MEDICAL LONGITUDINAL SERIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Forecasting how 3D medical scans evolve over time is important for disease progression, treatment planning, and developmental assessment. Yet existing models either rely on a single prior scan, fixed grid times, or target global labels, which limits voxel-level forecasting under irregular sampling. We present CRONOS, a unified framework for many-to-one prediction from multiple past scans that supports both discrete (grid-based) and continuous (real-valued) timestamps in one model, to the best of our knowledge the first to achieve continuous sequence-to-image forecasting for 3D medical data. CRONOS learns a spatio-temporal velocity field that transports context volumes toward a target volume at an arbitrary time, while operating directly in 3D voxel space. Across three public datasets spanning Cine-MRI, perfusion CT, and longitudinal MRI, CRONOS outperforms other baselines, while remaining computationally competitive. We will release code and evaluation protocols to enable reproducible, multi-dataset benchmarking of multi-context, continuous-time forecasting.

1 INTRODUCTION



(a) Task description.

(b) Memory Efficiency vs. Performance

Figure 1: **Task and benchmark comparison (a) Task setup** Forecasting a target 3D scan from multiple past volumes in two regimes. *Discrete:* acquisitions lie approximately on a regular grid, but may contain missing frames (dotted boxes). *Continuous:* acquisitions occur at irregular, real-valued timestamps and are used directly without grid alignment. Many-to-one task $(\{I_i\}_{i=1}^{T}, t_{\text{target}}) \rightarrow I_{\text{target}}$. **(b) Efficiency and performance** Left: GPU memory scaling of single forward pass with sequence length T shows CRONOS to be substantially more memory-efficient than alternatives. Right: Average SSIM across two datasets, where CRONOS outperforms baselines and LCI.

Longitudinal medical imaging is central to monitoring disease progression, assessing treatment response, and modeling anatomical development across time (Suter et al., 2022; Rivail et al., 2019; Bernard et al., 2018). Some modalities are inherently spatio-temporal, such as ultrasound (US), cine-MRI, videos, or perfusion Computer Tomography (CT). Beyond these, repeated clinical acquisitions form temporal sequences that may span over months or years and are used for clinical decision making. In ophthalmology, for instance, longitudinal OCT volumes are central to monitoring progression of age-related macular degeneration and predicting treatment response (Rivail et al., 2019). Works such as using surgical video streams (Li et al., 2024), which are also increasingly leveraged for diverse tasks, or in (Gomes et al., 2022), where longitudinal US sequences are used, show the overall breadth of spatio-temporal imaging. Beyond individual modalities, there is also a massive and growing amount of video and longitudinal data across clinical contexts (Farhad et al., 2023), including applications such as treatment response prediction in oncology (Suter et al., 2022).

Despite its importance, spatio-temporal learning in medical imaging is centered mostly on single time-point (image-to-image) analysis. Some approaches rely on global labels e.g. Yoon et al. (2024), while many reduce to image-to-image preidction with a single context scan (Zhang et al., 2025a)). Ohters introduce task-specific prior or remain tied to one disease (e.g. Puglisi et al. (2025). In particular, Alzheimer's Disease (AD) has attracted a disproportionate share of longitudinal imaging research (Petersen et al., 2010; Martí-Juan et al., 2020; Chen et al., 2025), whereas other domains remain comparatively underexplored.

Category	**Method**	**C1**	**C2**	**C3**	**C4**
Med. Gen	BrLP	✗	✓	✓	✓
	LociDiffCom	✗	✓	✓	✗
	ImageFlowNet	✗	✓	✗	✓
Vid. Gen	MCVD	✓	✓	✗	✗
STL	SimVP	✓	✗	✓	✗
	ViViT	✓	✗	✓	✗
	ConvLSTM	✓	✗	✓	✗
	NODE+LSTM	✓	✗	✓	✓
Med. STL	CRONOS (ours)	✓	✓	✓	✓

Table 1: **Technical comparison of spatio-temporal prediction methods.** Columns denote Challenges ($C\#$): **C1: Multiple Inputs**, **C2: high fidelity**, **C3: 3D imaging**, **C4: continuous-time modeling**. Our proposed CRONOS satisfies all four criteria, whereas existing medical and natural imaging baselines lack one or more. STL stands for spatio-temporal learning.

CRONOS addresses these challenges by introducing a unified spatio-temporal flow framework for medical sequence-to-image prediction that: [1]

- **Supports both *discrete* and *continuous* timestamps**, leveraging multiple past scans jointly on **3D** medical imaging data.
- **Avoids disease-specific assumptions**, enabling application to any medical longitudinal task.
- **Consistently outperforms prior approaches**, including standard sequence models and the Last Context Image (LCI) baseline, which is a surprisingly simple and competitive heuristic (NRMSE, PSNR, and SSIM), due to slowly changing medical images.

2 Related Work

Medical Imaging Prior work in longitudinal medical imaging focuses heavily on one-to-one, or one-to-many video prediction. While approaches like diffusion models (Litrico et al., 2024; Zhu et al.; Puglisi et al., 2025) and Neural ODEs (Lachinov et al., 2022; Liu et al., 2025) have been applied to medical imaging, these are *image-to-image*, and thus cannot canonically capture multi-input longitudinal evolution. For example, Bai & Hong (2024) propose a continuous-time model, but they predict sequences from single images. In contrast, works that jointly leverage multiple observations show improved prediction accuracy (Fang et al., 2021). The single-context nature makes these aforementioned works not sufficient for our setting. There are also **interpolation-based methods** (Zhu et al., 2024) which predict intermediate frames between two acquisitions, but this restricts their use to filling missing intervals rather than forecasting. Overall, existing medical approaches are all technically restricted; be it only single-image input, disease specific priors, limited to 2D, or not being able to forecast to arbitrary times as shown in 1.

[1] Code will be released at `github.com/anonymous`.

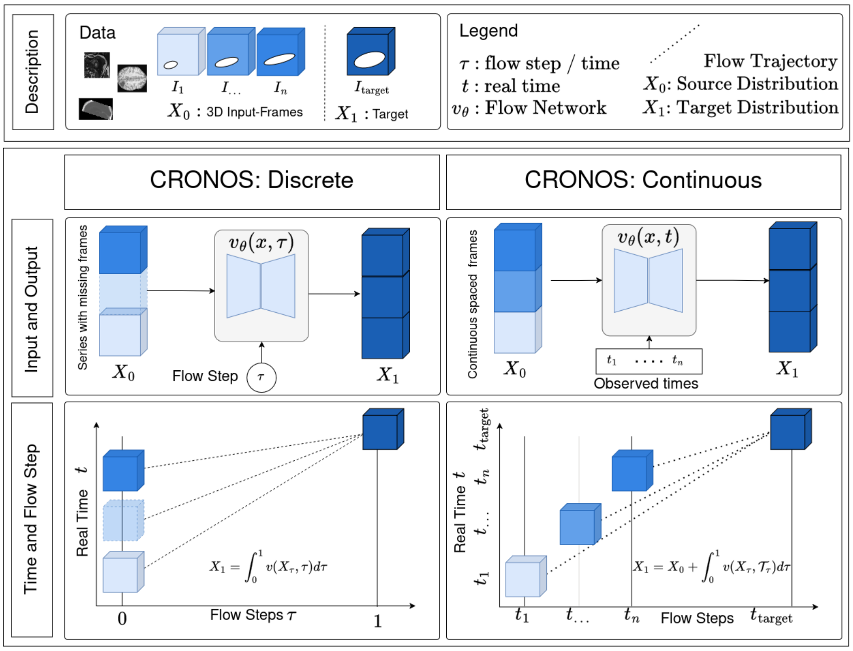


Figure 2: **CRONOS method overview: Left:** Discrete CRONOS treats time implicitly, interpolating between context frames and a fixed target along a normalized flow step $t \in [0, 1]$. **Right:** Continuous CRONOS explicitly conditions on real-valued timestamps t_i, allowing each context I_i to transport toward the target via its own interpolation t_i. This enables predictions at arbitrary target times while preserving the true temporal geometry.

By contrast, our work focuses on continuous-time modeling across full spatio-temporal sequences without restricting to specific modalities or diseases.

Natural Imaging and Video Prediction Spatio-temporal modeling has been extensively studied in video prediction. Early approaches such as ConvLSTM (SHI et al., 2015) introduced recurrent sequence-to-sequence architectures and remain widely used. Subsequent methods such as SimVP (Gao et al., 2022) replaced recurrence with purely convolutional designs. Transformer-based models like ViViT (Arnab et al., 2021) extended attention mechanisms to the video domain and have become a backbone in many imaging domains. More recent efforts have explored generative modeling, including video diffusion (Voleti et al., 2022; Ye & Bilodeau, 2023; Yan et al., 2021), and continuous-time formulations such as Neural ODEs (Chen et al., 2019), extended to videos in (Park et al., 2021). While these approaches are powerful, they have primarily been developed for dense 2D natural video sequences with large-scale training data. Accordingly, they transfer poorly to 3D medical images with small datasets and sparse sequences , thus motivating our work.

Flow Matching Flow Matching (FM) has recently emerged as a generative modeling paradigm (Lipman et al., 2023; 2024), and has been adapted to irregular time series, e.g. in (Zhang et al., 2025b), though only for low-dimensional data rather than full image sequences. Our extension therefore is: while classical FM learn a single flow from (most often) raw noise $X_0 \sim p$ to samples $X_1 \sim 1$ along steps $\tau \in [0, 1]$, we re-cast

$$X_0 = [I_1, \dots, I_T], \qquad X_1 = \mathcal{I}_{\text{target}} := [I_{\text{target}}, \dots, I_{\text{target}}], \tag{1}$$

interpreting p as the context sequence, and q as a broadcast stack of I_{target} (defined the stack as $\mathcal{I}_{\text{target}}$, to make dimension explicit). This temporal broadcasting turns FM into sequence-to-image transport: a shared velocity field v_θ simultaneously moves all T context volumes toward the target, effectively T per-frame transports under shared parameters. We refer to this framework as **Continuous RecOnstructioNs for medical lOngitudinal Series (CRONOS)**.

3 Methods

Algorithm 1 CRONOS Continuous: Training and Inference

```
Require: Patients $\mathcal{P}$ and initial network $v_\theta$
while training do
    Sample $\{[\mathcal{I}, I_{\text{target}}],\ [t_1, \ldots, t_T, t_{\text{target}}]\} \sim \mathcal{P}(\mathcal{X})$    ▷ pick a random patient
    Sample $\tau \sim \mathcal{U}(0, 1)$    ▷ random flow step
    $\mathcal{I}_{\text{target}} \leftarrow [I_{\text{target}}, \ldots, I_{\text{target}}]$    ▷ repeat target $T$ times
    $\mathcal{T}'_\tau \leftarrow (1-\tau)\,[t_1, \ldots, t_n] + \tau\, \boldsymbol{t}_{\text{target}}$    ▷ interpolate timestamps
    $X_\tau \leftarrow (1-\tau)\,\mathcal{I} + \tau\,\mathcal{I}_{\text{target}} + \sigma(\tau)\,\epsilon$    ▷ linear interpolation
    $\mathcal{L} \leftarrow \|v_\theta(\mathcal{T}'_\tau, X_\tau) - (\mathcal{I}_{\text{target}} - \mathcal{I})\|^2$    ▷ velocity loss
    Update $\theta \leftarrow \mathrm{AdamW}(\nabla_\theta \mathcal{L})$
return $v_\theta$
if inference then
    Initialize $X_0 \leftarrow \mathcal{I}$
    Define integration grid $\{\tau_0 = 0, \ldots, \tau_N = 1\}$ with $N$ steps
    $\mathcal{T}'_\tau = (1-\tau)\,[t_1, \ldots, t_n] + \tau\, \boldsymbol{t}_{\text{target}}$
    $\hat{X}_{0:N} \leftarrow \mathrm{ODEInt}(v_\theta, X_0, \{\mathcal{T}'_0, \ldots, \mathcal{T}'_1\})$    ▷ numerical integration
    return $\hat{X}_N$
```

3.1 Problem Setup

Let $\mathcal{P} = \left\{\, (\{I_i^{(n)}, t_i^{(n)}\}_{i=1}^{T^{(n)}},\ t_{\text{target}}^{(n)},\ I_{\text{target}}^{(n)})\, \right\}_{n=1}^{p}$ denote a dataset of p patient sequences. Each (patient) sequence consists of a set of T context volumes $\mathcal{I} = \{I_1, \ldots, I_T\}$, with $I_i \in \mathbb{R}^{H \times D \times W}$ (for shorthand $S = H \times D \times W$), acquired at associated timestamps $\{t_1, \ldots, t_T\} \subset \mathbb{R}_+$. We consider two regimes. ***Discrete***: Acquisitions lie on a uniform time grid; some frames may be missing, yielding sparse sequences (e.g., natural video, cine-MRI, perfusion CT). ***Continuous***: Acquisitions occur at irregular, real-valued times that do not easily align to any grid (typical in longitudinal clinical scans). For continuous series, forcing a frame grid either explodes sequence length with empty slots or loses temporal precision. For instance, daily-resolution for timepoints over several years would yield T in the thousands, yet in practice only a handful of scans are ever acquired. In both discrete and continuous series, T is small relative to natural video.

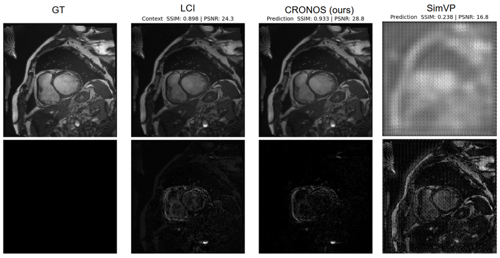

GT
LCI
Context SSIM: 0.898 | PSNR: 24.3
CRONOS (ours)
Prediction SSIM: 0.933 | PSNR: 28.8
SimVP
Prediction SSIM: 0.238 | PSNR: 16.8

Figure 3: **Qualitative comparison on the ACDC dataset**. Ground truth (GT), Last Context Image (LCI), our method (CRONOS), and SimVP. Upper row: prediction, lower row: residuals.

Target Task. Given the set of context images and time$\{(I_i, t_i)\}_{i=1}^{T}$, as well as a target time t_{target}, we aim to learn

$$f\big(\{I_i, t_i\}_{i=1}^{T},\ t_{\text{target}}\big) \mapsto I_{\text{target}}. \tag{2}$$

The discrete setting uses a fixed grid (with optional zero-tensors for missing context volumes); the continuous setting uses the **observed** context only, without padding.

3.2 Flow Matching (FM)

Flow Matching Lipman et al. (2023) learns a ordinary differential equation (ODE), linking the equal dimensional distributions p and q via

$$\frac{d}{d\tau}\psi_\tau(x) = u_\tau(\psi_\tau(x)), \qquad X_1 = X_0 + \int_0^1 u_\tau(X_\tau)\, d\tau, \tag{3}$$

with $X_0 \sim p$, $X_1 \sim q$. A convenient coupling is obtained by sampling X_τ as

$$X_\tau = (1-\tau)X_0 + \tau X_1 + \sigma(\tau)\epsilon, \tag{4}$$

where $\epsilon \sim \mathcal{N}(0, I)$ denotes random gaussian noise and $\sigma(\tau)$ its intensity, which is sampled around the straight path. The corresponding ground-truth velocity along this path is therefore constant:

$$u_\tau(X_\tau) = \frac{d}{d\tau}X_\tau = X_1 - X_0. \tag{5}$$

Consequently, to approximate the ground truth velocity, we train a neural network $v_\theta(X_\tau, \tau) \in \mathbb{R}^{T \times S}$ using:

$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{X_0, X_1, \tau} \big\| v_\theta(X_\tau, \tau) - u_\tau(X_\tau) \big\|_2^2. \tag{6}$$

Using v_θ, we can then infer using equation 3 via an approximate ODE solver.

3.3 Continuous and discrete reconstructions for medical image time series (CRONOS)

We introduce CRONOS, a spatio-temporal flow model that learns continuous trajectories from longitudinal scans. It comes in two complementary variants: *discrete* and *continuous*.

Temporal broadcasting for sequence-to-image flows To enable flow between a sequence of context images and a single target, we define $X_0 \sim p$ as the stack of context images (with variant-specific handling for continuous vs. discrete), and $X_1 \sim q$ as the target image broadcast to the same shape

$$X_1 = [I_{\text{target}}, \dots, I_{\text{target}}]. \tag{7}$$

This broadcasting ensures that X_0 and X_1 share the same dimensionality, allowing us to define a valid flow between them.

Discrete CRONOS. On a regular grid with missing scans, we first *embed* each sequence onto the grid of a resolution g using a binning operator $\mathcal{E}^{\text{grid}}_{\mathbf{g}}$, which assigns each I_i to the closes grid index matching t_i (proper definition in A.1.1). Missing slots are then handled by a last-observed carry-forward operator $\mathcal{F}^{\text{LOCF}}$, which fills empty positions with the most recent available scan. In short, we define

$$X_0 = \big(\underbrace{\mathcal{F}^{\text{LOCF}}}_{\text{fill}} \circ \underbrace{\mathcal{E}^{\text{grid}}_{\mathbf{g}}}_{\text{bin to grid}}\big)\big(\{(I_i, t_i)\}_{i=1}^{T}\big) = [\hat{I}_1, \dots, \hat{I}_K]. \tag{8}$$

This pre-processing ensures X_0 is well-defined on a uniform grid. Furthermore, LOCF handles spatial missingness: missing frames are zero-initialized and replaced by the most recent observation (Appendix A.1.2). This setup stabilizes optimization and preserves grid order while enabling many-to-one sequence transport within FM. Finally, we train on the linear interpolation $X_\tau = (1-\tau)X_0 + \tau X_1$ using equation 6, where temporal order is captured *implicitly* by the flow step τ and the frame index. Additionally, we set $\sigma = 0$ during training and inference, ablations on nonzero noise levels are reported in Table 7.

Continuous CRONOS Our continuous modeling strategy extends on the discrete case by conditioning on *real-valued timestamps* while evolving along a scalar flow parameter $\tau \in [0, 1]$. We construct spatio-temporal tensors (using mild abuse of notation): Time enters the network only as conditioning on real timestamps. We interpolate the conditioning timestamps along the interpolated time vector $\mathcal{T}_\tau$. X_0 is then defined as in equation equation 7, without embedding it to the grid, nor performing LOCF as in discrete CRONOS. We define the shifted time vector as

$$\mathcal{T}_\tau = (1-\tau)\,\mathbf{t}_{\text{ctx}} + \tau\,\mathbf{t}_{\text{target}}. \tag{9}$$

The formulation in equation 9 lets flow step τ carry real temporal information, without adding extra complexity. The conditional trajectory is then

$$X_1 \;=\; X_0 + \int_0^1 v_\theta\big(X_\tau, \mathcal{T}_\tau\big)\, d\tau, \tag{10}$$

where v_θ is the predicted velocity field and τ *is the flow step* (usually called time, we avoid it due to avoiding confusion). Prediction is then done via approximate solution of equation 3, solver details found in C.1. This formulation lets CRONOS model continuous image evolution grounded in actual scan times, supporting interpolation or forecasting without regular sampling or artificial frame filling. It avoids zero-padding, leading to reduced computational burden compared to the discrete variant. Both variants use the same 3D U-Net backbone, further details are provided in Appendix C.3, and the training/inference procedure appears in Algorithm 1.

Time Encoding. Flow steps and continuous times are mapped to Fourier embeddings using (Tancik et al., 2020), which were used e.g. in (Rombach et al., 2022): $\gamma(t) = [\sin(2\pi f_k t), \cos(2\pi f_k t)]_{k=1}^{K}$ using frequencies f_k. To preserve dimensional consistency across variable-length input sequences for the continuous setting, we compute the time embedding as

$$\text{Enc}(\boldsymbol{t}) = \frac{1}{T} \sum_{i=1}^{T} \gamma(t_i). \tag{11}$$

This embedding is then added to each residual layer via FiLM. The loss is then calculated via equation 6, and inference via equation 10.

4 Data and Experimental Design

4.1 Datasets

ACDC (Bernard et al., 2018) is a cardiac MRI dataset capturing different heart phases. The context tensor is reshaped to $[T, H, D, W] = [11, 32, 128, 128]$, and the target is a single image with the same spatial size. We split ACDC into 80 training, 20 validation and 50 test images. This dataset served for method development; ablations were conducted on the validation split.
ISLES (Riedel et al., 2024) consists of perfusion CT image time series from stroke patients. From the normalized series, we sample 7 consecutive points, take the last as the target, and randomly mask the remaining context frames. The resulting context tensor has shape $[T, H, D, W] = [7, 16, 128, 128]$. We use a split of 92 training, 23 validation and 34 test images. For both the ACDC and dataset, we randomly mask out time points (see Appendix C.2).
Lumiere (Suter et al., 2022) is a longitudinal glioma MRI dataset with 3D scans. Images are reshaped to $[T, H, D, W] = [7, 96, 96, 64]$. Because some patients have few acquisitions, we prepend zeros to standardize pre-processing across cases. The split is 48 training, 12 validation and 14 test images.

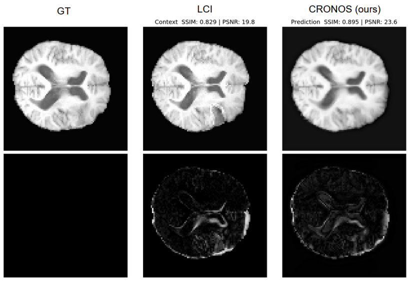

GT
LCI
Context SSIM: 0.829 | PSNR: 19.8
CRONOS (ours)
Prediction SSIM: 0.895 | PSNR: 23.6

Figure 4: **Qualitative comparison on the LUMIERE dataset**. Ground truth (GT), Last Context Image (LCI), our method (CRONOS), and SimVP baseline. Lumiere is particularly challenging due the very small dataset. highlighting the benefit of explicit continuous-time conditioning under extreme data scarcity.

4.2 Experimental settings

Reproducibility details can be found in Section C.

Discrete Setting: As mentioned in the data section, input data has dimension T, while some frames may be missing. We apply *both* variants of CRONOS, noting that the continuous version can also operate in this regime with a smaller context window, since missing images *do not need* to be explicitly represented. The lower context window also leads to a lower computational demand. Therefore, the underlying tensors remain uniform, with some time points masked. For validation and testing we ensure that the missingness pattern is fixed across epochs, as otherwise the choice of best checkpoint would be ill-posed (further details in Appendix C.2).

Continuous Setting: As an *additional ablation and experiment*, we simulate a continuous setup on ACDC to highlight the gains from explicit timestamp conditioning. While no public dataset provides plenty of continuous acquisition protocols, this sub-sampled variant shows that CRONOS benefits from real-valued time even beyond irregular masking. Specific details of how we subsampled ACDC can be found in C.1. Importantly, both the discrete and continuous formulations remain applicable to discrete grids.

Table 2: **Discrete Time: Quantitative Evaluation on Many-to-One Sequences:** Reported values are mean (standard deviation) over three runs. Metrics include normalized root MSE, $NRMSE$, structural similarity index ($SSIM[\%]$) and peak signal-to-noise-ratio $PSNR$. *ViViT OOM on a 40 GB GPU, despite having a smaller batch size and the lowest possible feature size. Standard deviation of LCI omitted for visual clarity. Blue row: only method to beat LCI and our proposed CRONOS. Computational requirements on ACDC in A

Dataset	Model	NRMSE [10^{-2}] ↓	SSIM [%] ↑	PSNR [dB] ↑
ACDC	LCI	4.48	92.79	28.918
	ConvLSTM	11.20 ± 0.48	50.44 ± 1.53	19.123 ± 0.312
	SimVP	9.27 ± 0.29	49.08 ± 4.01	20.715 ± 0.267
	NODE + LSTM	11.59 ± 0.18	36.41 ± 2.94	18.946 ± 0.186
	ViViT	13.90 ± 2.66	17.06 ± 8.60	17.252 ± 1.738
	CRONOS discrete	3.97 ± 1.23	**94.51 ± 0.79**	**30.510 ± 1.560**
	CRONOS cont.	**3.74 ± 0.21**	94.34 ± 0.45	29.750 ± 0.528
ISLES	LCI	5.25	96.29	29.002
	ConvLSTM	19.31 ± 0.18	39.92 ± 0.66	17.644 ± 0.014
	SimVP	13.06 ± 0.19	48.82 ± 1.60	20.799 ± 0.112
	ViViT	16.54 ± 0.30	36.76 ± 1.49	18.671 ± 0.134
	NODE + LSTM	15.10 ± 0.87	40.55 ± 7.15	19.481 ± 0.515
	CRONOS discrete	4.50 ± 0.76	**97.33 ± 0.93**	30.542 ± 1.540
	CRONOS cont.	**4.38 ± 0.48**	97.31 ± 0.38	**30.809 ± 1.099**
Lumiere	LCI	8.38	88.35	21.631
	ConvLSTM	34.79 ± 0.67	9.21 ± 2.81	9.217 ± 0.171
	SimVP	71.03 ± 0.89	-1.92 ± 0.51	2.989 ± 0.109
	ViViT*	OOM	OOM	OOM
	NODE+LSTM	13.07 ± 1.03	48.66 ± 2.26	17.742 ± 0.659
	CRONOS discrete	7.92 ± 0.92	**91.43 ± 1.84**	22.427 ± 0.969
	CRONOS cont.	**7.55 ± 0.86**	89.32 ± 1.83	**22.551 ± 0.979**

Baselines We compare CRONOS against established spatio-temporal learning methods. As a clinically motivated heuristic, the Last Context Image baseline (LCI) simply reuses the last available image and serves as a lower bound. Among sequence models, we include ConvLSTM (SHI et al., 2015), SimVP (Gao et al., 2022), and ViViT (Arnab et al., 2021) as representative recurrent, convolutional, and transformer backbones. For continuous-time sequence modeling, we further evaluate an ODE-LSTM (Lechner & Hasani, 2020) baseline. For the flow matching library we use Tong et al. (2024b;a); Tong (2025). Together, these methods provide a spectrum of spatio-temporal architectures against which we benchmark CRONOS. Computational requirements are described in detail in the appendix.

Continuous vs. Discrete. We report results in two regimes: an *discrete* setting, which allows direct comparison to existing spatio-temporal baselines, and a *continuous* setting on ACDC, designed as an ablation to test the benefit of explicit timestamp conditioning.

5 Results and Discussion

5.1 Towards unified benchmarking for medical 3D sequence-to-image forecasting

We are among the first to propose an experimental setup for the sequence-to-image task, evaluating CRONOS under two complementary regimes. The first uses *discrete* input sequences, where some context images are missing but but acquisitions lie on a regular grid. This setting enables comparison against established spatio-temporal baselines. The second uses ACDC with resampled acquisitions to mimic *continuous* input, allowing us to assess the benefit of explicit timestamp conditioning.

For completeness, we include an image-to-image (*not sequence-to-image*) diffusion baseline on ACDC (details in B.5) This required a two-stage training setup, first pretraining an autoencoder and then training the diffusion module for 1000 denoising steps, which already made the approach far more computationally demanding than all other baselines. Iterative denoising leads to an order-of-magnitude longer inference time for a single image-to-image step and several orders of magnitude higher training cost, while not surpassing the simple LCI heuristic. [2]

5.2 CRONOS IS STATE-OF-THE-ART FOR SPATIO-TEMPORAL 3D MEDICAL IMAGE FORECASTING

Table 2 reports the quantitative results across all three datasets. We observe that both variants of CRONOS **substantially outperform** the competing spatio-temporal baselines, as well as LCI. We also note that individually, CRONOS is better than LCI on each individual validation run. On LUMIERE, which is characterized by very sparse and heterogeneous tumor trajectories, it is surprising that CRONOS is even able to outperform LCI. These results demonstrate that CRONOS is effective across different temporal regimes: the discrete formulation already yields strong performance, while the continuous formulation provides further gains when timestampts are informative. CRONOS runs **within the same computational budget** during inference (see Figure 1b) and in similar orders of magnitude (VRAM and wall-clock time) during training as natural imaging baselines (see 8). Further ablations are provided in A, confirming that CRONOS is stable across variations in *feature size, training noise, and integration settings*. While small differences appear, they are not substantial, indicating that our network is *highly robust* to hyperparameter choices.

Method	SSIM ↑	PSNR ↑	NRMSE ↓
LCI	93.27	29.77	0.0349
NODE + LSTM	57.50	22.87	0.0728
CRONOS discr.	93.27	29.77	0.0348
CRONOS cont.	93.86	30.09	0.0330

Table 3: **Continuous ACDC**, where discrete CRONOS lacks explicit timestamp conditioning, and therefore fails to outperform LCI. Additional experiments in B and in Table 9

5.3 CRONOS ENABLES EFFICIENT FLOW-BASED CONTINUOUS MEDICAL MODELING

Table 3 demonstrates that *incorporating explicit time embeddings improves forecasting* quality when scans occur at irregular intervals. This shows that the continuous formulation of CRONOS is not only feasible but also beneficial in realistic clinical settings, where images are often irregularly sampled. In fact, if we fully remove the timestamp information entirely, performance differences increase significantly, and the continuous variant clearly outperforms the discrete one 9. Together, these results highlight that modeling real-valued timestamps can provide a measurable advantage over treating sequences as grid-aligned. However, in Table 2, we see that using the discrete variant remains highly competitive. Although any irregular series can in principle be quantized to a grid via $\mathcal{E}_{\mathbf{g}}^{\text{grid}}$, doing so without loss requires increasingly fine grids. This becomes computationally inefficient, whereas the continuous variant scales with the number of context images and *not* with the grid range $K \cdot \Delta$ (see equation 12). This is reflected in Table 8 and Figure 1b, where continuous CRONOS is both more memory-efficient and faster to train than the discrete formulation. It also highlights a broader limitation of the field: the scarcity of diverse spatio-temporal datasets in which real timing information is critical.

5.4 CRONOS PRODUCES SHARPER RECONSTRUCTIONS WIT LOWER RESIDUALS

Figures 3, 6 and 4 highlight qualitative comparisons, as well as dataset examples. The LCI baseline often appears visually close to the target, largely because many longitudinal scans exhibit only subtle changes. However, e.g. SimVP tends to introduce artifacts and blur anatomical details. In contrast, CRONOS yields sharper reconstructions and consistently lower residuals compared to LCI, highlighting its ability to capture fine-grained temporal progression.

[2] On our setup, a naive auto-regressive image-to-image *latent-diffusion* pipeline applied across 11 context times per subject requires ∼5–6 hours *per validation step*; see Appendix A for details.

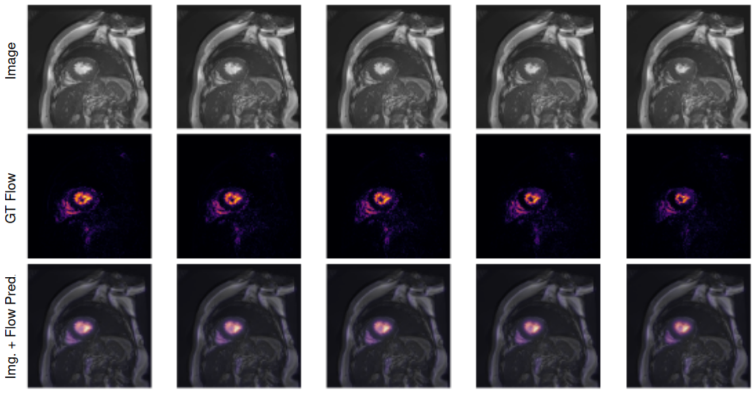


Figure 5: **Network Flows**: Top: input images at the first five timestamps. Middle: ground-truth voxel-wise differences ($|I_i - I_{\text{target}}|$. Bottom: predicted velocity fields $v_\theta(X_0, 0)$, overlaid on the corresponding inputs. The highlighted regions coincide with the areas of the largest temporal changes (primarily the ventricular cavities and myocardial boundaries).

5.5 Future work: unlocking general spatio-temporal medical forecasting

While voxel-wise fidelity metrics such as NRMSE, PSNR, and SSIM remain the community standard, they do not fully capture clinically relevant trajectory modeling. As highlighted in recent efforts on image analysis validation Maier-Hein et al. (2024), such metrics may not always align with actual domain interest. Developing metrics for spatio-temporal forecasting is therefore an important future direction. In parallel, the scarcity of longitudinal and spatio-temporal datasets (beyond the ones we used in this study), poses a broader challenge for robust evaluation. Encouragingly, our results on LUMIERE suggest that progress is possible even under severe data limitations, and we hope to motivate further work on curating larger and more diverse publicly available cohorts. Finally, the absence of large-scale foundation models for medical imaging, particularly in the spatio-temporal domain, remains a major bottleneck. We view our work as a keystone contribution: establishing a unified flow-based framework for continuous spatio-temporal medical volumetric forecasting that can *both benefit from, and motivate*, future developments in medical imaging.

6 Conclusion

In this work, we presented CRONOS (Continuous RecOnstructioNs for medical lOngitudinal Series), a unified spatio-temporal framework that forecasts 3D medical volumes at arbitrary target times by combining multiple context scans with explicit real-valued time conditioning. Unlike single-image or time-agnostic methods, CRONOS handles both grid-aligned and continuous timestamps within one architecture, and makes no disease-specific assumptions, it is among the first methods to demonstrate continuous sequence-to-image forecasting for 4D medical data. Across three publicly available datasets (Cine-MRI, perfusion CT, longitudinal MRI), it outperforms baselines-including the strong Last Context Image

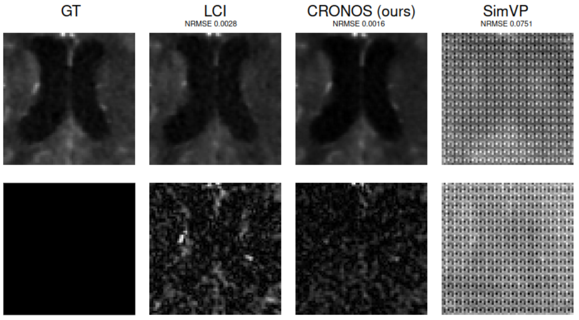


Figure 6: **Zoomed-in qualitative comparison on the ISLES dataset**. Ground truth (GT), Last Context Image (LCI), our method (CRONOS), and SimVP baseline. Shown here for visibility is a zoomed in patch of the qualitative results of the ISLES dataset.

(LCI)-and remains robust under hyper-parameter changes while remaining computationally competitive. By resolving the aforementioned limitations, our method enables clinically specific studies and advances patient-level forecasting for personalized precision medicine.

Broader Impact

Longitudinal modeling of medical images has the potential to improve patient care by enabling earlier detection of disease progression, monitoring of treatment response, and improved personalization of therapy. By explicitly modeling continuous temporal evolution, our approach could support clinicians in making more informed decisions. However, there are also risks: mispredictions may lead to incorrect clinical conclusions if models are deployed without careful validation and without a human in the loop. Biases in training data (e.g., underrepresentation of certain populations or imaging modalities) may propagate to predictions, raising concerns about fairness and generalizability, which is a common problem in medical imaging. We emphasize that our method is a research contribution intended to advance especially technical methodology. Clinical deployment would require extensive validation, regulatory approval, and integration into existing workflows. We believe that by releasing code and benchmarks, this work will support the community in building transparent, reproducible, and safe spatio-temporal models for healthcare. But by proposing this method, we hope to support a general-purpose foundation for medical spatio-temporal and longitudinal modeling, which could massively propel this area forward.

References

Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. ViViT: A Video Vision Transformer. pp. 6836–6846, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Arnab_ViViT_A_Video_Vision_Transformer_ICCV_2021_paper.html?ref=https://githubhelp.com.

Hao Bai and Yi Hong. NODER: Image Sequence Regression Based on Neural Ordinary Differential Equations, July 2024. URL http://arxiv.org/abs/2407.13241. arXiv:2407.13241 [cs].

Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, Gerard Sanroma, Sandy Napel, Steffen Petersen, Georgios Tziritas, Elias Grinias, Mahendra Khened, Varghese Alex Kollerathu, Ganapathy Krishnamurthi, Marc-Michel Rohé, Xavier Pennec, Maxime Sermesant, Fabian Isensee, Paul Jäger, Klaus H. Maier-Hein, Peter M. Full, Ivo Wolf, Sandy Engelhardt, Christian F. Baumgartner, Lisa M. Koch, Jelmer M. Wolterink, Ivana Išgum, Yeonggul Jang, Yoonmi Hong, Jay Patravali, Shubham Jain, Olivier Humbert, and Pierre-Marc Jodoin. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525, November 2018. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2018.2837502. URL https://ieeexplore.ieee.org/document/8360453/.

Durong Chen, Meiling Zhang, Hongjuan Han, Yalu Wen, and Hongmei Yu. Reflections on dynamic prediction of Alzheimer's disease: advancements in modeling longitudinal outcomes and time-to-event data. *BMC Medical Research Methodology*, 25(1):175, July 2025. ISSN 1471-2288. doi: 10.1186/s12874-025-02618-x. URL https://doi.org/10.1186/s12874-025-02618-x.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural Ordinary Differential Equations, December 2019. URL http://arxiv.org/abs/1806.07366. arXiv:1806.07366 [cs].

Cong Fang, Song Bai, Qianlan Chen, Yu Zhou, Liming Xia, Lixin Qin, Shi Gong, Xudong Xie, Chunhua Zhou, Dandan Tu, Changzheng Zhang, Xiaowu Liu, Weiwei Chen, Xiang Bai, and Philip H. S. Torr. Deep learning for predicting COVID-19 malignant progression. *Medical Image Analysis*, 72:102096, August 2021. ISSN 1361-8415. doi: 10.1016/j.media.2021.102096. URL https://www.sciencedirect.com/science/article/pii/S1361841521001420.

Moomal Farhad, Mohammad Mehedy Masud, Azam Beg, Amir Ahmad, and Luai Ahmed. A Review of Medical Diagnostic Video Analysis Using Deep Learning Techniques. *Applied Sciences*, 13(11):6582, January 2023. ISSN 2076-3417. doi: 10.3390/app13116582. URL https://www.mdpi.com/2076-3417/13/11/6582. Publisher: Multidisciplinary Digital Publishing Institute.

Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z. Li. SimVP: Simpler Yet Better Video Prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3170–3180, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Gao_SimVP_Simpler_Yet_Better_Video_Prediction_CVPR_2022_paper.html.

Ryan G. Gomes, Bellington Vwalika, Chace Lee, Angelica Willis, Marcin Sieniek, Joan T. Price, Christina Chen, Margaret P. Kasaro, James A. Taylor, Elizabeth M. Stringer, Scott Mayer McKinney, Ntazana Sindano, George E. Dahl, William Goodnight III, Justin Gilmer, Benjamin H. Chi, Charles Lau, Terry Spitz, T. Saensuksopa, Kris Liu, Jonny Wong, Rory Pilgrim, Akib Uddin, Greg Corrado, Lily Peng, Katherine Chou, Daniel Tse, Jeffrey S. A. Stringer, and Shravya Shetty. AI system for fetal ultrasound in low-resource settings, March 2022. URL http://arxiv.org/abs/2203.10139. arXiv:2203.10139 [cs].

Dmitrii Lachinov, Arunava Chakravarty, Christoph Grechenig, Ursula Schmidt-Erfurth, and Hrvoje Bogunovic. Learning Spatio-Temporal Model of Disease Progression with NeuralODEs from

Longitudinal Volumetric Data, November 2022. URL http://arxiv.org/abs/2211.04234. arXiv:2211.04234 [cs].

Mathias Lechner and Ramin M. Hasani. Learning Long-Term Dependencies in Irregularly-Sampled Time Series. *ArXiv*, June 2020. URL https://www.semanticscholar.org/paper/Learning-Long-Term-Dependencies-in-Time-Series-Lechner-Hasani/4d9521fbd135559e4d186e96b703f3bd8fd7617e.

Yunlong Li, Zijian Zhao, Renbo Li, and Feng Li. Deep learning for surgical workflow analysis: a survey of progresses, limitations, and trends. *Artificial Intelligence Review*, 57(11): 291, September 2024. ISSN 1573-7462. doi: 10.1007/s10462-024-10929-6. URL https://doi.org/10.1007/s10462-024-10929-6.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling, February 2023. URL http://arxiv.org/abs/2210.02747. arXiv:2210.02747 [cs].

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow Matching Guide and Code, December 2024. URL http://arxiv.org/abs/2412.06264. arXiv:2412.06264 [cs].

Mattia Litrico, Francesco Guarnera, Mario Valerio Giuffrida, Daniele Ravì, and Sebastiano Battiato. TADM: Temporally-Aware Diffusion Model for Neurodegenerative Progression on Brain MRI. In Marius George Linguraru, Qi Dou, Aasa Feragen, Stamatia Giannarou, Ben Glocker, Karim Lekadir, and Julia A. Schnabel (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume 15002, pp. 444–453. Springer Nature Switzerland, Cham, 2024. ISBN 978-3-031-72068-0 978-3-031-72069-7. doi: 10.1007/978-3-031-72069-7_42. URL https://link.springer.com/10.1007/978-3-031-72069-7_42. Series Title: Lecture Notes in Computer Science.

Chen Liu, Ke Xu, Liangbo L. Shen, Guillaume Huguet, Zilong Wang, Alexander Tong, Danilo Bzdok, Jay Stewart, Jay C. Wang, Lucian V. Del Priore, and Smita Krishnaswamy. ImageFlowNet: Forecasting Multiscale Image-Level Trajectories of Disease Progression with Irregularly-Sampled Longitudinal Medical Images, April 2025. URL http://arxiv.org/abs/2406.14794. arXiv:2406.14794 [eess].

Lena Maier-Hein, Annika Reinke, Patrick Godau, Minu D. Tizabi, Florian Buettner, Evangelia Christodoulou, Ben Glocker, Fabian Isensee, Jens Kleesiek, Michal Kozubek, Mauricio Reyes, Michael A. Riegler, Manuel Wiesenfarth, A. Emre Kavur, Carole H. Sudre, Michael Baumgartner, Matthias Eisenmann, Doreen Heckmann-Nötzel, Tim Rädsch, Laura Acion, Michela Antonelli, Tal Arbel, Spyridon Bakas, Arriel Benis, Matthew B. Blaschko, M. Jorge Cardoso, Veronika Cheplygina, Beth A. Cimini, Gary S. Collins, Keyvan Farahani, Luciana Ferrer, Adrian Galdran, Bram van Ginneken, Robert Haase, Daniel A. Hashimoto, Michael M. Hoffman, Merel Huisman, Pierre Jannin, Charles E. Kahn, Dagmar Kainmueller, Bernhard Kainz, Alexandros Karargyris, Alan Karthikesalingam, Florian Kofler, Annette Kopp-Schneider, Anna Kreshuk, Tahsin Kurc, Bennett A. Landman, Geert Litjens, Amin Madani, Klaus Maier-Hein, Anne L. Martel, Peter Mattson, Erik Meijering, Bjoern Menze, Karel G. M. Moons, Henning Müller, Brennan Nichyporuk, Felix Nickel, Jens Petersen, Nasir Rajpoot, Nicola Rieke, Julio Saez-Rodriguez, Clara I. Sánchez, Shravya Shetty, Maarten van Smeden, Ronald M. Summers, Abdel A. Taha, Aleksei Tiulpin, Sotirios A. Tsaftaris, Ben Van Calster, Gaël Varoquaux, and Paul F. Jäger. Metrics reloaded: recommendations for image analysis validation. *Nature Methods*, 21(2):195–212, February 2024. ISSN 1548-7105. doi: 10.1038/s41592-023-02151-z. URL https://www.nature.com/articles/s41592-023-02151-z. Publisher: Nature Publishing Group.

Gerard Martí-Juan, Gerard Sanroma-Guell, and Gemma Piella. A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease. *Computer Methods and Programs in Biomedicine*, 189:105348, June 2020. ISSN 0169-2607. doi: 10.1016/j.cmpb.2020.105348. URL https://www.sciencedirect.com/science/article/pii/S0169260719316165.

Sunghyun Park, Kangyeol Kim, Junsoo Lee, Jaegul Choo, Joonseok Lee, Sookyung Kim, and Edward Choi. Vid-ODE: Continuous-Time Video Generation with Neural Ordinary Differential Equation, March 2021. URL http://arxiv.org/abs/2010.08188. arXiv:2010.08188 [cs].

R C. Petersen, P S. Aisen, L A. Beckett, M C. Donohue, A C. Gamst, D J. Harvey, C R. Jack, W J. Jagust, L M. Shaw, A W. Toga, J Q. Trojanowski, and M W. Weiner. Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neurology*, 74(3):201–209, January 2010. ISSN 0028-3878. doi: 10.1212/WNL.0b013e3181cb3e25. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2809036/.

Lemuel Puglisi, Daniel C. Alexander, and Daniele Ravì. Brain Latent Progression: Individual-based spatiotemporal disease progression on 3D Brain MRIs via latent diffusion. *Medical Image Analysis*, pp. 103734, July 2025. ISSN 1361-8415. doi: 10.1016/j.media.2025.103734. URL https://www.sciencedirect.com/science/article/pii/S1361841525002816.

Evamaria O. Riedel, Ezequiel de la Rosa, The Anh Baran, Moritz Hernandez Petzsche, Hakim Baazaoui, Kaiyuan Yang, David Robben, Joaquin Oscar Seia, Roland Wiest, Mauricio Reyes, Ruisheng Su, Claus Zimmer, Tobias Boeckh-Behrens, Maria Berndt, Bjoern Menze, Benedikt Wiestler, Susanne Wegener, and Jan S. Kirschke. ISLES 2024: The first longitudinal multimodal multi-center real-world dataset in (sub-)acute stroke. 2024. doi: 10.48550/ARXIV.2408.11142. URL https://arxiv.org/abs/2408.11142. Publisher: arXiv Version Number: 1.

Antoine Rivail, Ursula Schmidt-Erfurth, Wolf-Dieter Vogl, Sebastian M. Waldstein, Sophie Riedl, Christoph Grechenig, Zhichao Wu, and Hrvoje Bogunović. Modeling Disease Progression In Retinal OCTs With Longitudinal Self-Supervised Learning, October 2019. URL http://arxiv.org/abs/1910.09420. arXiv:1910.09420 [eess].

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685, June 2022. doi: 10.1109/CVPR52688.2022.01042. URL https://ieeexplore.ieee.org/document/9878449/. Conference Name: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) ISBN: 9781665469463 Place: New Orleans, LA, USA Publisher: IEEE.

Xingjian SHI, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html.

Yannick Suter, Urspeter Knecht, Waldo Valenzuela, Michelle Notter, Ekkehard Hewer, Philippe Schucht, Roland Wiest, and Mauricio Reyes. The LUMIERE dataset: Longitudinal Glioblastoma MRI with expert RANO evaluation. *Scientific Data*, 9(1):768, December 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01881-7. URL https://www.nature.com/articles/s41597-022-01881-7. Publisher: Nature Publishing Group.

Matthew Tancik, Pratul P. Srinivasan, B. Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, R. Ramamoorthi, J. Barron, and Ren Ng. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *ArXiv*, June 2020. URL https://www.semanticscholar.org/paper/Fourier-Features-Let-Networks-Learn-High-Frequency-Tancik-Srinivasan/a0dc3135c40e150f0271002a96b7c9680b6cac40.

Patrick Therasse, Susan G. Arbuck, Elizabeth A. Eisenhauer, Jantien Wanders, Richard S. Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T. Van Oosterom, Michaele C. Christian, and Steve G. Gwyther. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. *JNCI: Journal of the National Cancer Institute*, 92(3):205–216, February 2000. ISSN 0027-8874, 1460-2105. doi: 10.1093/jnci/92.3.205. URL https://academic.oup.com/jnci/jnci/article/2965042/New. Publisher: Oxford University Press (OUP).

Alexander Tong. TorchCFM, January 2025. URL https://github.com/atong01/conditional-flow-matching.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport, March 2024a. URL http://arxiv.org/abs/2302.00482. arXiv:2302.00482 [cs].

Alexander Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching, March 2024b. URL http://arxiv.org/abs/2307.03672. arXiv:2307.03672 [cs].

Vikram Voleti, Alexia Jolicoeur-Martineau, and Christopher Pal. MCVD: Masked Conditional Video Diffusion for Prediction, Generation, and Interpolation, October 2022. URL http://arxiv.org/abs/2205.09853. arXiv:2205.09853 [cs].

Wilson Yan, Yunzhi Zhang, P. Abbeel, and A. Srinivas. VideoGPT: Video Generation using VQ-VAE and Transformers. *ArXiv*, April 2021. URL https://www.semanticscholar.org/paper/VideoGPT%3A-Video-Generation-using-VQ-VAE-and-Yan-Zhang/2d9ae4c167510ed78803735fc57ea67c3cc55a35.

Xi Ye and Guillaume-Alexandre Bilodeau. STDiff: Spatio-temporal Diffusion for Continuous Stochastic Video Prediction. 2023. doi: 10.48550/ARXIV.2312.06486. URL https://arxiv.org/abs/2312.06486. Publisher: arXiv Version Number: 1.

Dan Yoon, Youho Myong, Young Gyun Kim, Yongsik Sim, Minwoo Cho, Byung-Mo Oh, and Sungwan Kim. Latent diffusion model-based MRI superresolution enhances mild cognitive impairment prognostication and Alzheimer's disease classification. *NeuroImage*, 296:120663, August 2024. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2024.120663. URL https://www.sciencedirect.com/science/article/pii/S1053811924001587.

Shaorong Zhang, Tamoghna Chattopadhyay, Sophia I. Thomopoulos, Jose-Luis Ambite, Paul M. Thompson, and Greg Ver Steeg. Diffusion Bridge Models for 3D Medical Image Translation, April 2025a. URL http://arxiv.org/abs/2504.15267. arXiv:2504.15267 [cs].

Xi Zhang, Yuan Pu, Yuki Kawamura, Andrew Loza, Yoshua Bengio, Dennis L. Shung, and Alexander Tong. Trajectory Flow Matching with Applications to Clinical Time Series Modeling, February 2025b. URL http://arxiv.org/abs/2410.21154. arXiv:2410.21154 [cs].

Zihao Zhu, Tianli Tao, Yitian Tao, Haowen Deng, Xinyi Cai, Gaofeng Wu, Kaidong Wang, Haifeng Tang, Lixuan Zhu, Zhuoyang Gu, Dinggang Shen, and Han Zhang. LoCI-DiCom: Longitudinal Consistency-Informed Diusion Model for 3D Infant Brain Image Completion.

Zihao Zhu, Tianli Tao, Yitian Tao, Haowen Deng, Xinyi Cai, Gaofeng Wu, Kaidong Wang, Haifeng Tang, Lixuan Zhu, Zhuoyang Gu, Dinggang Shen, and Han Zhang. LoCI-DiffCom: Longitudinal Consistency-Informed Diffusion Model for 3D Infant Brain Image Completion. In Marius George Linguraru, Qi Dou, Aasa Feragen, Stamatia Giannarou, Ben Glocker, Karim Lekadir, and Julia A. Schnabel (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume 15002, pp. 249–258. Springer Nature Switzerland, Cham, 2024. ISBN 978-3-031-72068-0 978-3-031-72069-7. doi: 10.1007/978-3-031-72069-7_24. URL https://link.springer.com/10.1007/978-3-031-72069-7_24. Series Title: Lecture Notes in Computer Science.