PATTERN-GUIDED DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models have shown promise in forecasting future data from multivariate time series. However, few existing methods account for recurring structures, or patterns, that appear within the data. We present Pattern-Guided Diffusion Models (PGDM), which leverage inherent patterns within temporal data for forecasting future time steps. PGDM first extracts patterns using archetypal analysis and estimates the most likely next pattern in the sequence. By guiding predictions with this pattern estimate, PGDM makes more realistic predictions that fit within the set of known patterns. We additionally introduce a novel uncertainty quantification technique based on archetypal analysis, and we dynamically scale the guidance level based on the pattern estimate uncertainty. We apply our method to two well-motivated forecasting applications, predicting visual field measurements and motion capture frames. On both, we show that pattern guidance reduces PGDM's prediction error by up to 40.67% and 11.10%, respectively. Compared to baselines, PGDM also achieves lower error by up to 65.58% and 82.54%.

1 INTRODUCTION

Diffusion models are a class of generative models that perform generation by iteratively removing noise from a noisy sample. These models are easier to train and generate higher quality images compared to the previous state-of-the-art, generative adversarial networks (Dhariwal & Nichol, 2021). Recent work has found success in using diffusion models to forecast future steps of temporal data (Chang et al., 2024; Feng et al., 2024; Gu et al., 2022; Hu et al., 2024; Li et al., 2022; Lv et al., 2024; Rasul et al., 2021; Wen et al., 2023). Such methods, however, rarely leverage the recurring structures that often manifest in temporal data. These appear, for example, in medical modalities due to the physiology and anatomy of the human body. Basketball videos also contain repeated structures due to standardized courts, player positions, and strategies. The few diffusion-based forecasters that exploit these *patterns* often overlook changes over time and uncertainties in pattern representation (Wang et al., 2024; Westny et al., 2024; Zhao et al., 2024).

In this paper, we present Pattern-Guided Diffusion Models (PGDM) for forecasting temporal data with inherent patterns. Using archetypal analysis (Cutler & Breiman, 1994), we extract patterns from training data, then train a guidance function to predict future pattern contributions to the data. PGDM then forecasts future points guided by these predictions. To handle evolving patterns, we introduce a novel uncertainty metric that dynamically tunes the scale of pattern guidance.

We evaluate PGDM on two impactful applications. First, we consider the clinical application of visual field prediction. Visual field tests measure a patient's functional vision, and the resulting measurements manifeset common patterns across patients due to the anatomy of the eye. Furthermore, forecasting future visual field measurements can serve as a decision aid for clinicians. On a real-world visual field dataset, we find that pattern guidance reduces the error of PGDM predictions by up to 40.67% on average. Furthermore, PGDM achieves up to 65.58% lower error on average compared to baseline models. Next, we consider the application of forecasting future motion capture frames for human motion prediction. Pose patterns frequently appear in common human movements, such as walking and running. Predicting human motion may aid advancements in human robot collaboration and autonomous driving. On motion capture frames for a variety of dance genres, we show that PGDM is able to leverage even the diverse, highly dynamic patterns that present in dance motion. Pattern guidance reduces the error of PGDM predictions by up to 11.10% on average, allowing PGDM to surpass baselines by up to 82.54% on average.

In summary, our contributions are as follows.

1. We present Pattern-Guided Diffusion Models (PGDM), which leverage inherent patterns within temporal data.
2. We introduce a novel uncertainty quantification method based on archetypal analysis, and we show that this uncertainty metric captures geometric distance from the training set.
3. We show that the proposed uncertainty quantification metric approximately lower bounds the error of the pattern predictions that guide PGDM.
4. We propose a method to dynamically tune the level of pattern guidance based on the proposed uncertainty metric.

2 Related Works

Diffusion models have been widely applied to temporal forecasting across many modalities and applications. TimeGrad (Rasul et al., 2021) and LDT (Feng et al., 2024) forecast multivariate time series conditioning on historical data, while BVAE (Li et al., 2022) uses a bi-directional VAE for the reverse diffusion process. Models like USTD (Hu et al., 2024) and DiffSTG (Wen et al., 2023) focus on spatio-temporal graphs, capturing spatial dependencies. Other applications include pedestrian trajectory (Lv et al., 2024; Gu et al., 2022) and medical sensor signal prediction (Chang et al., 2024). For a comprehensive overview, see Yang et al. (2024).

Few such diffusion-based forecasters attempt to leverage patterns within the data. Hypothesizing that past patterns tend to reappear later, Diff-MGR (Zhao et al., 2024) conditions predictions on previous patterns. Westny et al. (2024) also proposed to guide predictions of traffic trajectories using patterns. As agent behaviors are often dictated by the environment (e.g., cars are likely to stay within the lanes of a road), this approach conditions predictions on a map of the environment. Similarly to Diff-MGR, Westny et al. (2024) assumes that apparent patterns within the temporal data will remain constant over time, as predictions are conditioned on a fixed map. Wang et al. (2024) instead proposed to guide predictions with dynamically changing patterns, captured by real-time camera readings. They noted that humans are most likely to walk towards specific destinations within a scene, such as a door, stairs, or a hallway. Their proposed EgoNav diffusion model therefore used segmented image data as conditioning inputs for predicting human walking trajectories. However, these approaches do not account for uncertainty in the guiding patterns. In contrast, our PGDM model adapts the level of pattern guidance based on the estimated reliability of dynamically evolving patterns.

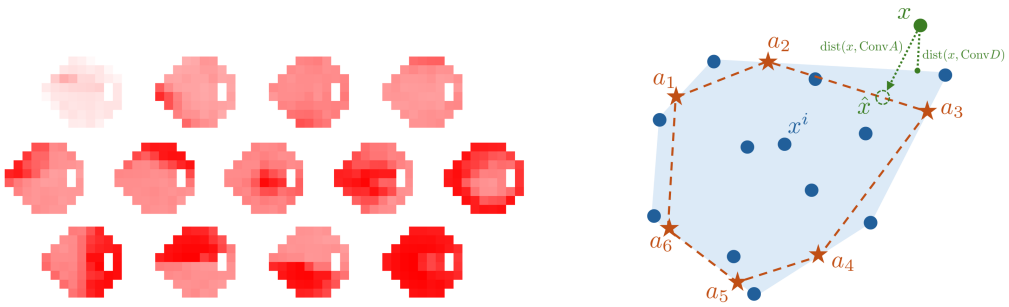
3 Background

Let the data of interest be $x \in \mathbb{R}^d$ sampled from distribution $p(x)$, which arrives in a temporal sequence $\{x_t\}$ with time index t. We are concerned with such data that contains *patterns*, or repeating structures. Given an observed history of length T over time $t \in \{1, 2, \dots, T\}$, we aim to predict a horizon of length H over time $t \in \{T+1, T+2, \dots, T'\}$ with $T' = T + H$. Denote a set of n of history and horizon pairs by $\{x^i_{1:T}, x^i_{T:T'}\}_{i=1}^n$. We use archetypal analysis, which identifies patterns resembling real data rather than an abstraction, to overcome the challenge of extracting useful patterns. Here, we provide a brief overview of diffusion models and archetypal analysis.

3.1 Diffusion Models

Diffusion models (Ho et al., 2020; Sohl-Dickstein et al., 2015) aim to learn and generate data from a distribution $p(x)$ through a forward and reverse process, in which noise is iteratively added to and removed from the data, respectively. Given $x_{t,0} \sim p(x_t)$ at time t, the fixed S-step forward process creates a sequence of increasingly noisy samples $x_{t,1}, x_{t,2}, \dots, x_{t,S}$. Note that here we use the notation $x_{t,s}$, where t denotes the time index and s denotes the diffusion step. The noisy samples are drawn from the distribution $q\left(x_{t,s}|x_{t,s-1}\right) \coloneqq \mathcal{N}\left(\sqrt{1-\beta_s}x_{t,s-1}, \beta_s\mathbf{I}, \right)$, where $\beta_1, \beta_2, \dots, \beta_S$ is a noise variance schedule. With appropriately chosen variance schedule, this distribution approaches a standard normal as $S \to \infty$. Conveniently, the noising step $x_{t,s}$ can be sampled in closed form given $x_{t,0}$, $q\left(x_{t,s}|x_{t,0}\right) = \mathcal{N}\left(\sqrt{\bar{\alpha}_s}x_{t,0}, (1-\bar{\alpha}_s)\mathbf{I}\right)$, where $\alpha_s = 1 - \beta_s$ and $\bar{\alpha}_s = \prod_{i=1}^{s} \alpha_i$.

Conversely, the reverse process removes noise from $x_{t,S} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ to ultimately recover the training distribution. The goal is to learn the distribution $p_\theta\left(x_{t,s-1}|x_{t,s}\right) \coloneqq \mathcal{N}\left(\mu_\theta\left(x_{t,s}, s\right), \sigma_s^2\mathbf{I}\right)$. The



(a) Example visual field archetypes.

(b) Geometric interpretation.

Figure 1: Overview of archetypal analysis (AA). (a) Example archetypes extracted from a visual field dataset. The archetypes capture visual loss patterns that consistently appear across glaucoma patients. Darker regions indicate greater vision loss. (b) Given a dataset D (blue dots), AA identifies a set of archetypes A (red stars). Within $\text{Conv}A \subseteq \text{Conv}D$ (within red dashed line), any point can be reconstructed from A without error. For any point $x \notin \text{Conv}A$, the reconstruction error is the distance between x and $\text{Conv}A$.

denoising parameters θ are learned by optimizing the evidence lower bound (ELBO) on negative log likelihood, $-\log p_\theta\left(x_{t,0}\right)$. At inference time, samples from the learned distribution $p_\theta\left(x_{t,0}\right)$ are generated by applying the reverse process over S steps to noisy samples.

3.2 Archetypal Analysis

Archetypal analysis (AA) (Cutler & Breiman, 1994) extracts extremal patterns, or archetypes, from a given dataset. These archetypes are themselves a combination of the data and therefore are realistic and interpretable representations of the data's significant patterns. Figure 1a shows examples extracted from a visual field dataset. Furthermore, any point within the given dataset can be constructed as a combination of these archetypes, allowing for the contribution of each pattern to be quantified.

More formally, given dataset $D = \{x^i\}_{i=1}^n$, AA finds the p archetypes $a_1, a_2, \ldots a_p \in \mathbb{R}^d$ that minimize the residual sum of squares (RSS) error,

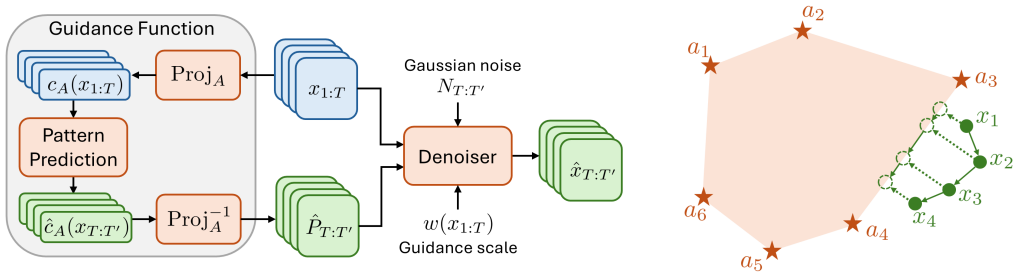
$$\min_{c_j^i, a_j^i \ \forall i,j} \sum_{i=1}^n \left\| x^i - \sum_{j=1}^p c_j^i a_j \right\|^2, \tag{1}$$

subject to $c_j^i \geq 0 \ \forall j$ and $\sum_{j=1}^p c_j^i = 1$ for each $i \in \{1, 2, \ldots, n\}$, and where $a_j = \sum_{k=1}^n \beta_j^k x^k$ with $\beta_j^k \geq 0 \ \forall k$ and $\sum_{k=1}^n \beta_j^k = 1$ for each $j \in \{1, , 2, \ldots, p\}$. That is, if zero reconstruction error is achieved, then each $x^i \in D$ can be reconstructed as a convex combination of the archetypes, and the archetypes are themselves a convex combination of the data. Then, given a set of identified archetypes, any new data point can be reconstructed as a convex combination of the archetypes, with coefficients found by minimizing the objective in equation 1 with fixed $a_1, a_2, \ldots, a_p$.

Figure 1b visualizes a geometric interpretation of AA. Intuitively, AA identifies a region in which any point can be perfectly reconstructed by the archetypes. This region is the convex hull of the archetypes, denoted $\text{Conv}A$ for the archetype set A. In the ideal case when $p = n$, the set of archetypes is exactly D, and the region of reconstructible points fully encompasses the dataset. In practice, $p < n$ is typically chosen, and the resulting archetypes instead lie on the boundary of $\text{Conv}D$ (Cutler & Breiman, 1994, Proposition 1). The archetype set therefore defines a reconstructible region that closely, but often not fully, captures the dataset D. Furthermore, for any $x \notin \text{Conv}A$, the reconstruction error can be thought of as the distance between x and $\text{Conv}A$. In Section 4.2, we show that this reconstruction error can be used to estimate the distance between x and the dataset D.

4 Methods

Our goal is to learn the conditional distribution $p(x_{T:T'}|x_{1:T}, \hat{P}_{T:T'})$, where $\hat{P}_{T,T'}$ represents the patterns estimated to manifest in the future. To determine a pattern representation space, we first extract archetype set A from the training data. The representation space is the contribution of



(a) PGDM.

(b) AA information loss.

Figure 2: (a) Overview of Pattern-Guided Diffusion Models. A pattern guidance function estimates the future patterns over a horizon. This prediction is performed in the archetype space, where the projected representation captures the contribution of each pattern to the data. The predicted future patterns are supplied as conditioning context to the diffusion model. The scale of the pattern guidance is dynamically tuned based on the uncertainty of the input sequence. (b) Projection into the lower-dimensional archetype space loses information for any point outside the convex hull of the archetypes $\mathrm{Conv}A$. In this example, the sequence $x_1, x_2, x_3, x_4 \notin \mathrm{Conv}A$ is perturbed to the boundary of $\mathrm{Conv}A$ after reconstruction, losing important dynamical information.

archetypes to each data point. We then train our proposed Pattern-Guided Diffusion Model (PGDM) to learn the desired conditional distribution.

Figure 2a summarizes PGDM, which is constructed from three key components. 1) A pattern guidance function predicts the patterns $\hat{P}$ that will appear over the future horizon. First, the input sequence $x_{1:T}$ is projected to the archetype space, where the projected pattern representation $c_A(x_{1:T})$ quantifies the contribution of each pattern to the data. The projected representation for the future steps is then predicted, and the resulting estimate $\hat{c}_A(x_{T:T'})$ is lifted back to the original data space as a predicted pattern $\hat{P}_{T:T'}$. 2) While prediction in the lower p-dimensional archetype space helps to evade the curse of dimensionality, the projection operation naturally incurs an information loss. By design, while any point $x_t \in \mathrm{Conv}A$ can be projected with no information loss, this cannot be said for a point $x_t \notin \mathrm{Conv}A$, which may appear when the data's patterns change over time. Figure 2b visualizes this loss of information. We therefore introduce a novel Archetypal Analysis Uncertainty Quantification (AAUQ) technique, which we show captures geometric distance from the training set. This uncertainty metric also estimates a lower bound on the loss of the guidance function. 3) We use the predicted pattern from the guidance function as additional conditioning context for the diffusion model, following the general methodology of classifier-free diffusion guidance (Ho & Salimans, 2022). To account for uncertainties in the guidance function, we dynamically tune the guidance scale based on our uncertainty metric. We now describe each of the three components in further detail.

4.1 Guidance Function

First, we employ archetypal analysis to extract significant patterns from the training data. For convenience, when the meaning is clear, we overload the notation A to indicate both the resulting set of $p < d$ archetypes $\{a_1, a_2, \ldots, a_p\} \subset \mathbb{R}^d$ and the matrix constructed from these archetypes $A \in \mathbb{R}^{d \times p}$. These archetypes define an archetype, or pattern, space in p dimensions. Given archetypes A, let the projection function that determines the pattern representation (i.e., the coefficients minimizing objective equation 1) be $c_A : \mathbb{R}^d \to \mathbb{R}^p$. For any point x_t, the reconstruction is $\hat{x}_t = Ac_A(x_t)$. We denote the error of this reconstruction as $L_{c_A}(x_t) = \|x_t - Ac_A(x_t)\|$.

Next, we train a lightweight neural network $f_A : \mathbb{R}^{p \times T} \to \mathbb{R}^{p \times H}$ to predict H future pattern representations based on T past pattern representations. The error of this prediction function is $L_{f_A}(c_{1:T}) = \|Ac_{T:T'} - Af_A(c_{1:T})\|$.

The guidance function $f_G : \mathbb{R}^{d \times T} \to \mathbb{R}^{d \times H}$ is

$$f_G(x_{1:T}) = Af_A \circ c_A(x_{1:T}). \tag{2}$$

Note that $f_G(x_{1:T})$ is $\hat{P}_{T:T'}$ in Figure 2a. We now show that a lower bound on the guidance function error L_{f_G} is a function of the projection error L_{c_A} and pattern prediction error L_{f_A}.

Theorem 1 (Bound on Guidance Function Error)**.** *For any sequence $x_{1:T}$ and horizon $x_{T:T'}$ pair, let the error of guidance function f_G defined in Equation equation 2 be $L_{f_G}(x_{1:T}) = \|x_{T:T'} - Af_A(c_A(x_{1:T}))\|$. Then*

$$L_{f_G}(x_{1:T}) \geq L_{c_A}(x_{T:T'}) - L_{f_A}(x_{1:T}). \tag{3}$$

Proof. Please see Appendix A for the proof. □

It is clear from Theorem 3 that, if the prediction function f_A has a reasonably low error, then the error of the projection function c_A can serve as an approximate lower bound for the error of the guidance function f_G. Therefore, an estimate of L_{c_A} may quantify the degree to which PGDM should "trust" the guidance function, allowing the level of pattern guidance to be dynamically tuned. Next, we introduce a novel uncertainty quantification technique that can be used as a proxy for L_{c_A}.

4.2 Uncertainty Quantification for Archetypal Analysis

Projection to the lower-dimensional archetype space may lose information. We therefore introduce a novel uncertainty quantification metric based on archetypal analysis that captures this loss.

Definition 1 (Archetypal Analysis Uncertainty Quantification)**.** *For any sequence $x_{1:T}$ and archetype set A, the uncertainty u_A of the archetype projection is*

$$u_A(x_{1:T}) = \tfrac{1}{T}\sum_{t=1}^{T} \|x_t - Ac_A(x_t)\|. \tag{4}$$

Intuitively, this Archetypal Analysis Uncertainty Quantification (AAUQ) metric is simply the average reconstruction loss of the history sequence. AAUQ can also be geometrically interpreted as estimating the average distance of $x_{1:T}$ from the training dataset.

Theorem 2 (AAUQ as Geometric Distance)**.** *Assume that a set of archetypes $A = \{a_j\}_{j=1}^{p}$ is extracted from a dataset D. Define d as the closest point in $\mathrm{Conv}D$ to x_t. For any x_t,*

$$u_A(x_t) - \delta \leq \mathrm{dist}(x_t, \mathrm{Conv}D) \leq u_A(x_t) + \delta, \tag{5}$$

where $\delta = \|Ac_A(x_t) - d\|$ and $\delta = 0$ when $p = n$.

Proof. Please see Appendix B for the proof. □

Remark 1. *In Theorem 2, δ captures the ability of the archetypes to express the dataset D. This can be seen in the proof of Theorem 2. This can also be understood from the example in Figure 1b, in which $\mathrm{dist}(x, \mathrm{Conv}A)$ is exactly $u_A(x)$, and δ is the distance between $\hat{x}$ and the point on the boundary of $\mathrm{Conv}D$. This has interesting implications for the geometric interpretation of AAUQ in the case that the archetypes perfectly reconstruct the data, which occurs when the number of archetypes is equal to the size of the dataset. It is clear from Theorem 2 that if $p = n$,*

$$u_A(x_t) = \mathrm{dist}(x_t, \mathrm{Conv}D). \tag{6}$$

4.3 Pattern-Guided Diffusion Models

PGDM predicts future sequences $x_{T:T'}$ conditioned on the pattern prediction from the guidance function. We follow the methodology of classifier-free diffusion guidance, in which the model is trained with conditioning dropout. This effectively learns two denoising models, $\epsilon_\theta(z_{T:T',s},\ x_{1:T},\ \hat{P}_{T:T'})$ and $\epsilon_\theta(z_{T:T',s},\ x_{1:T},\ \emptyset)$, where $\emptyset$ is a null value and $z_{T:T',s}$ is the sample to be denoised at diffusion step s. Algorithm 1 summarizes the training process of PGDM. For history and horizon sequences sampled in Line 2, the guidance conditioning is set to the pattern prediction or a null value in Line 4. With the loss function on Line 5, the denoising model learns to estimate the noise added to $x_{T:,T'}$.

When generating samples with traditional classifier-free guidance, each denoising step uses a linear combination of the conditional and unconditional predictions:

$$\hat{\epsilon}_\theta(z_{T:T',s},\ x_{1:T}) = w\epsilon_\theta(z_{T:T',s},\ x_{1:T},\ \hat{P}_{T:T'}) + (1-w)\epsilon_\theta(z_{T:T',s},\ x_{1:T},\ \emptyset),$$

where $w \geq 0$ is the *guidance scale*. When $w = 0$, generation is unguided and sampled data are more diverse. The guidance level increases with w, leading to less diverse but higher quality samples.

Algorithm 1 PGDM Training

1: **for** all epochs **do**
2: Sample $x_{1:T}$, $x_{T:T'}$ from the training set
3: $s \sim \text{Uniform}(1, \dots, S)$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
4: $\hat{P}_{T:T'} = \emptyset$ with probability p_{drop}, else $\hat{P}_{T:T'} = f_G(x_{1:T})$ from eqn. equation 2
5: Take gradient descent step on $\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha}_s} x_{T:T'} + \sqrt{1-\overline{\alpha}_s}\epsilon,\ x_{1:T},\ \hat{P}_{T:T'})\|^2$
6: **end for**

Algorithm 2 PGDM Inference

1: **given** $x_{1:T}$
2: $x_{T:T',S} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
3: $\hat{P}_{T:T'} = f_G(x_{1:T})$ from eqn. equation 2
4: **for** s = S, ..., 1 **do**
5: $n \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if s > 1, else $n = 0$
6: Compute $\hat{\epsilon}_\theta(z_{T:T',s},\ x_{1:T})$ from eqn. equation 7
7: Sample $x_{T:T',s-1} = \frac{1}{\sqrt{\alpha_s}}\left(x_{T:T',s} - \frac{1-\alpha_s}{\sqrt{1-\overline{\alpha}_s}}\right)\hat{\epsilon}_\theta(z_{T:T',s},\ x_{1:T}) + \sqrt{\beta_s} n$
8: **end for**
9: Compute $w^* = w(x_{1:T}, 1, \gamma)$ from eqn. equation 8
10: Mix $\hat{x}_{T:T'} = w^* \hat{P}_{T:T'} + (1 - w^*) x_{T:T',0}$

While traditional classifier-free guidance is performed with constant guidance scale w, we instead use a dynamic guidance scale $w(x_{1:T})$ that captures the trustworthiness of the guidance function:

$$\hat{\epsilon}_\theta(z_{T:T',s},\ x_{1:T}) = w(x_{1:T})\epsilon_\theta(z_{T:T',s},\ x_{1:T},\ \hat{P}_{T:T'}) + (1 - w(x_{1:T}))\epsilon_\theta(z_{T:T',s},\ x_{1:T},\ \emptyset), \quad (7)$$

Definition 2 (Dynamic Guidance Scale)**.** *Let A be the set of archetypes extracted from the training dataset. Given a maximum guidance scale $\overline{w}$ and maximum tolerable uncertainty γ, the dynamic guidance scale for sequence $x_{1:T}$ is*

$$w(x_{1:T}, \overline{w}, \gamma) = \text{ReLU}\left(-\frac{\overline{w}}{\gamma} u_A(x_{1:T}) + \overline{w}\right). \quad (8)$$

Here, we measure the trustworthiness of our guidance function by our AAUQ uncertainty metric. The uncertainty $u_A(x_{1:T})$ is easy to compute and in practice, we find that the $u_A(x_{1:T})$ is a good proxy for estimating $L_{c_A}(x_{T:T'})$ and therefore the lower bound in Theorem 1. Hence, we design the dynamic guidance scale so that, as uncertainty increases, the guidance scale $w \in (0, \overline{w})$ decreases. When the uncertainty exceeds γ, w = 0. In other words, PGDM follows the pattern guidance most strictly when the data is in-distribution. For out-of-distribution data with unseen patterns, PGDM relies less on pattern-guidance, reverting to a standard diffusion model in the extreme case.

Algorithm 2 summarizes the inference process of PGDM. In Line 2, noisy data is sampled from a standard normal distribution. In Line 3, the patterns of the future sequence are predicted. Then, over S reverse diffusion steps in Lines 4–7, the guided and unguided denoising models are combined with dynamic guidance scale to iteratively remove noise. The resulting sequence prediction is sampled from the learned distribution, which we expect to be tightly centered around the pattern prediction. Finally in Lines 9 and 10, we mix the raw pattern prediction with the pattern-guided sequence prediction using a dynamic mixing scale. We include this mixing step to mitigate some potential practical challenges of PGDM. A full discussion of pattern mixing is included in Appendix C.

5 Applications

We validate PGDM on two applications, visual field prediction and human motion prediction. To aid in analyzing the effects of pattern guidance, we show the performance of two PGDM models selected from our hyperparameter search for both applications. The first model, PGDM_{MAE}, achieved the lowest validation mean absolute error (MAE) with pattern guidance. The second model, PGDM_{GDE}, achieved the highest capacity for pattern guidance, or the greatest achievable improvement in MAE by using guided predictions ($\overline{w} > 0$) over unguided predictions ($\overline{w} = 0$).

For these two models, we compare performance with guidance to performance without guidance and multiple baselines. While we would ideally compare PGDM to baselines that use some pattern conditioning (Wang et al., 2024; Westny et al., 2024; Zhao et al., 2024), most existing methods formalize patterns in a manner that is specific to the intended application, and therefore do not translate to our applications. For others, we were unable to obtain sufficient implementation details or code. Therefore, we instead select more general baseline techniques that do not use pattern conditioning. For both applications, we compare PGDM to two well-cited and common diffusion-based baselines for forecasting, TimeGrad (Rasul et al., 2021) and CSDI (Tashiro et al., 2021). While CSDI is a data imputation technique, it can easily be extended for forecasting. For the visual field application, we additionally compare PGDM to GenViT (Yang et al., 2022), a diffusion model with a vision transformer backbone that has been applied to VF prediction by Tian et al. (2023).

Source code is supplied in the supplementary material. Complete implementation details for pattern extraction and training, including hyperparameter selection, model architectures, and compute resources, are provided in Appendix D.

Visual Field Prediction. Pattern-Guided Diffusion models are especially useful in medical settings, where data often reflects consistent patterns due to anatomy. For instance, 24-2 visual field (VF) tests measure light sensitivity in decibels (dB) at 52 central points of vision, with specific loss patterns linked to structural eye damage (e.g., nerve fiber bundle loss) (Keltner et al., 2003). Figure 1a illustrates archetypal patterns from VF data, where darker areas indicate reduced vision. Forecasting VF outcomes can support clinicians in diagnosis, progression identification, and treatment planning.

We evaluate PGDM on the public the University of Washington Humphrey Visual Field (UWHVF) dataset (Montesano et al., 2022). To the best of our knowledge, UWHVF is the only publicly available 24-2 VF dataset, containing 7,428 sequences from 3,871 patients. The UWHVF measurements capture the patient's light sensitivity compared to normative data, ranging from -38 dB to 50 dB. That is, a negative (positive) dB indicates worse (better) vision than typical. Due to the few follow-up visits per patient, we predict $H = 1$ step into the future based on the past $T = 3$ steps. Additionally, as VF measurements are taken at non-constant time increments, we also condition predictions on the recorded age at each VF measurement and the desired time horizon for prediction. Thus, the one-step-ahead prediction can be made for an arbitrary length time period. We create multiple forecasting sequences from each patient in a sliding window fashion, resulting in 6,171 sequences.

Human Motion Prediction. We also apply PGDM to predict future frames of human motion capture data, a task relevant to domains including human robot interaction and autonomous driving (Lyu et al., 2022). Human motion often involves repeated body positions when executing common movements (e.g., walking, running, dancing). While our visual field application demonstrates PGDM's utility in the clinical domain, the motion prediction application presents a more challenging task with more rapidly evolving signals and longer prediction horizons.

We evaluate PGDM on the AIST++ dataset (Li et al., 2021), which contains motion capture frames capturing 3D motion from 10 dance genres. Predicting dance motion is more challenging for PGDM compared to locomotion, as the variety of dance genres and styles leads to a rich set of patterns with less periodic progressions over time. The AIST++ dataset captures the skeleton with 3D pose data for 17 keypoints, which represent specific joints or locations in the body. For consistency across heights, we normalize all data to the scale [0, 100]. We predict $H = 5$ steps into the future based on a past sequence of $T = 3$ steps. We create multiple forecasting sequences from each motion capture video, resulting between $36,739$ and $105,504$ sequences across the 10 genres.

5.1 Results

Pattern Extraction. We extract p = 13 archetypal patterns from UWHVF, shown in Figure 1a. We extract between p = 12 and p = 22 archetypes for each genre of AIST++. Figure 3 shows the archetypes extracted from the break dancing frames (see Appendix E for remaining genres). Appendix F reports the reconstruction error of the extracted patterns and the guidance function error.

AAUQ approximately lower bounds the guidance function error. Motivated by Theorem 1, PGDM uses AAUQ to determine the appropriate level of guidance. In Figure 4, we compare AAUQ measurements to the MAE of the guidance function. In both applications, we observe that AAUQ is indeed proportional to a linear lower bound on the guidance function error.

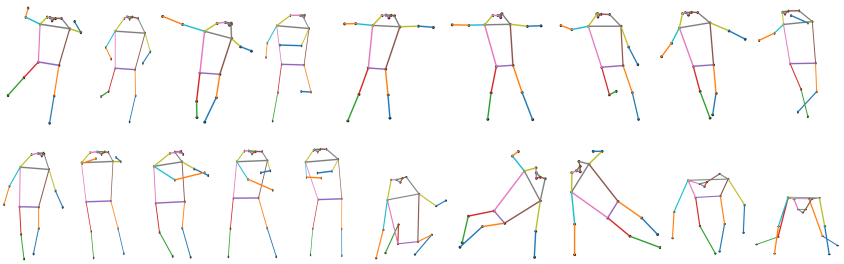
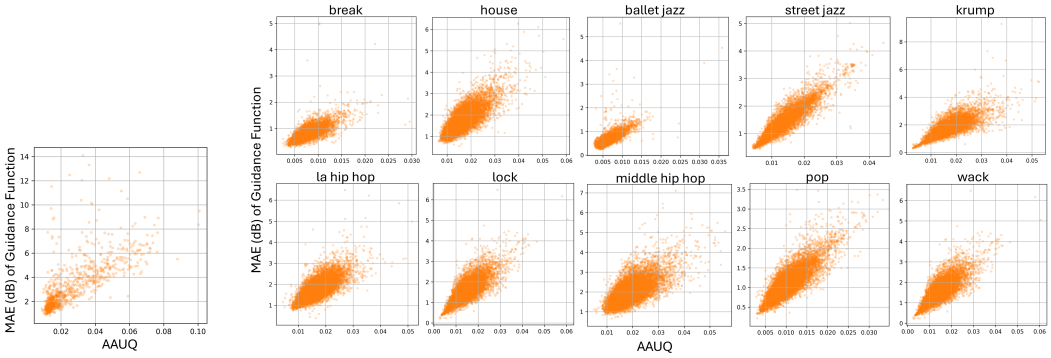


Figure 3: Nineteen archetypes extracted from AIST++ break dancing frames.



(a) VF prediction.

(b) Motion prediction.

Figure 4: AAUQ approximately lower bounds the guidance function error for both applications.

Pattern guidance reduces prediction error. Table 1 reports the mean absolute error (MAE) of PGDM_{MAE} and PGDM_{GDE} with and without pattern guidance. Compared to unguided predictions, pattern guidance reduces prediction error significantly. On UWHVF, guidance reduces the error of PGDM_{GDE} and PGDM_{MAE} by up to 40.67% and 20.90%, respectively. On AIST++, guidance reduces the error by up to 11.10% and 8.73%, respectively. To further study the impact of pattern guidance, we also show in Table 2 the performance of PGDM_{MAE} and PGDM_{GDE} on UWHVF with $\overline{w} = 1, 2, 3, 4, 5$. Similar results on AIST++ and qualitative examples are shown in Appendix G. In general, the standard deviation of MAE decreases as the guidance scale increases towards the optimal $\overline{w}$ choice, indicating that guidance improves both consistency and quality of predictions. Notably, we also find that excessive pattern guidance may lead to diminishing returns. In practice, an appropriate $\overline{w}$ may be selected in a manner similar to a hyperparameter search.

PGDM outperforms baselines. Table 1 also compares the performance of PGDM_{MAE} and PGDM_{GDE} to our baselines. Across the board, PGDM_{MAE} with pattern guidance achieves significantly lower MAE than GenViT, TimeGrad, and CSDI. On UWHVF, PGDM_{MAE} with guidance surpasses GenViT, TimeGrad, and CSDI by up to 65.58%, 29.36%, and 7.20%, respectively. On AIST++, PGDM_{MAE} surpasses TimeGrad and CSDI by up to 82.54% and 36.49%, respectively. However, we note that PGDM_{MAE} outperforms baselines even without guidance ($\overline{w} = 0$) on some AIST++ genres (e.g., break dancing). In these cases, to better demonstrate that PGDM's performance comes from pattern guidance, rather than model training alone, we also emphasize the results for PGDM_{GDE}. Even when unguided PGDM_{GDE} performs worse than baselines, pattern guidance almost always reduces the error of PGDM_{GDE} is to a lower or competitive level compared to baselines. These results demonstrate that pattern guidance is essential for higher quality predictions.

6 Conclusions

In this paper, we proposed Pattern-Guided Diffusion Models (PGDM), which leverage inherent archetypal patterns to forecast future steps from multivariate time series data. PGDM is guided by a pattern guidance function that predicts future patterns within the data. To estimate the trustworthiness of this guidance function, we introduced a novel uncertainty quantification metric that approximately

Table 1: Mean absolute error of PGDM_{MAE}, PGDM_{GDE}, and baselines for both the visual field prediction (UWHVF dataset) and the human motion prediction (10 dance genres from AIST++ dataset) case studies. For PGDM_{MAE} and PGDM_{GDE} with $\overline{w} > 0$, we show the guidance scale result that achieved the lowest error (for $\overline{w} = 1, 2, 3, 4, 5$ evaluations, see Appendix G). Mean and standard deviation are taken across five samples.

Visual Field Prediction						
GenViT	**TimeGrad**	**CSDI**	**PGDM_{GDE} $\overline{w} = 0$**	**PGDM_{GDE} $\overline{w} > 0$**	**PGDM_{MAE} $\overline{w} = 0$**	**PGDM_{MAE} $\overline{w} > 0$**
8.61±0.0018	4.19±0.0327	3.19±0.0421	5.20±0.0407	3.08±0.0153	3.75±0.0437	2.96±0.0117

Human Motion Prediction						
Genre	**TimeGrad**	**CSDI**	**PGDM_{GDE} $\overline{w} = 0$**	**PGDM_{GDE} $\overline{w} > 0$**	**PGDM_{MAE} $\overline{w} = 0$**	**PGDM_{MAE} $\overline{w} > 0$**
Break	2.10±0.0064	0.47±0.0010	0.52±0.0032	0.46±0.0013	0.41±0.0032	0.39±0.0011
House	3.71±0.0159	1.02±0.0045	0.90±0.0013	0.82±0.0017	0.79±0.0017	0.74±0.0011
Ballet Jazz	1.32±0.0098	0.55±0.0054	0.49±0.0004	0.45±0.0010	0.42±0.0008	0.39±0.0002
Street Jazz	1.65±0.0102	0.56±0.0054	0.60±0.0016	0.54±0.0005	0.52±0.0017	0.48±0.0008
Krump	2.37±0.0067	0.77±0.0017	0.88±0.0016	0.79±0.0011	0.77±0.0016	0.70±0.0013
LA Hip Hop	3.30±0.0157	0.78±0.0023	0.90±0.0009	0.82±0.0005	0.80±0.0010	0.74±0.0006
Lock	3.03±0.0086	0.76±0.0028	0.78±0.0017	0.71±0.0021	0.72±0.0011	0.67±0.0005
Middle Hip Hop	3.35±0.0113	1.04±0.0048	1.05±0.0034	0.96±0.0033	0.88±0.0014	0.82±0.0015
Pop	2.55±0.0105	0.70±0.0053	0.52±0.0013	0.49±0.0007	0.47±0.0008	0.44±0.0017
Wack	1.03±0.0047	0.44±0.0042	0.49±0.0054	0.47±0.0083	0.44±0.0044	0.41±0.0031

Table 2: Mean absolute error (MAE) of PGDM_{MAE} and PGDM_{GDE} on the VF prediction application (UWHVF dataset) with varying levels of guidance. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. GenViT**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
PGDM_{MAE}	$\overline{w} = 0$	3.75±0.0437	56.48±0.50	10.69±1.17	-17.32±1.55	-
	$\overline{w} = 1$	2.96±0.0117	65.58±0.14	29.36±0.67	7.20±1.25	20.90±0.71
	$\overline{w} = 2$	2.97±0.0112	65.54±0.13	29.29±0.67	7.10±1.27	20.81±0.73
	$\overline{w} = 3$	2.97±0.0108	65.51±0.13	29.23±0.67	7.03±1.29	20.75±0.76
	$\overline{w} = 4$	2.97±0.0107	65.48±0.13	29.17±0.67	6.95±1.31	20.68±0.78
	$\overline{w} = 5$	2.97±0.0104	65.45±0.12	29.10±0.66	6.86±1.32	20.60±0.80
PGDM_{GDE}	$\overline{w} = 0$	5.20±0.0407	39.60±0.47	-23.95±1.10	-62.83±2.59	-
	$\overline{w} = 1$	3.16±0.0212	63.28±0.24	24.65±0.76	1.01±1.44	39.21±0.23
	$\overline{w} = 2$	3.13±0.0201	63.67±0.23	25.44±0.76	2.06±1.43	39.85±0.23
	$\overline{w} = 3$	3.10±0.0187	63.93±0.21	25.99±0.74	2.77±1.41	40.29±0.24
	$\overline{w} = 4$	3.09±0.0170	64.09±0.20	26.32±0.72	3.20±1.39	40.55±0.25
	$\overline{w} = 5$	3.08±0.0153	64.16±0.18	26.47±0.70	3.40±1.39	40.67±0.27

lower bounds the guidance function error. Finally, we proposed to dynamically tune the level to which PGDM follows the pattern guidance based on this uncertainty metric. We found that PGDM outperforms baseline models, and pattern guidance reduces the error of PGDM. Two limitations of PGDM present interesting avenues for future work. First, PGDM has less benefit for out-of-distribution data exhibiting unseen patterns. Second, the use of AAUQ as an approximate lower-bound for guidance function error assumes that the temporal data is relatively continuous and does not rapidly change between the observed history and target prediction window. In some cases, this assumption is violated (e.g., rapidly changing signals sampled with a low frequency). Based on these limitations, PGDM may be further improved by updating the set of extracted patterns at inference time and accounting for signal dynamics when calculating the guidance scale.

REFERENCES

Ping Chang, Huayu Li, Stuart F Quan, Shuyang Lu, Shu-Fen Wung, Janet Roveda, and Ao Li. A transformer-based diffusion probabilistic model for heart rate and blood pressure forecasting in intensive care unit. *Computer Methods and Programs in Biomedicine*, 246:108060, 2024.

Adele Cutler and Leo Breiman. Archetypal analysis. *Technometrics*, 36(4):338–347, 1994.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Tobias Elze, Louis R Pasquale, Lucy Q Shen, Teresa C Chen, Janey L Wiggs, and Peter J Bex. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *Journal of The Royal Society Interface*, 12(103):20141118, 2015.

Shibo Feng, Chunyan Miao, Zhong Zhang, and Peilin Zhao. Latent diffusion transformer for probabilistic time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 11979–11987, 2024.

Tianpei Gu, Guangyi Chen, Junlong Li, Chunze Lin, Yongming Rao, Jie Zhou, and Jiwen Lu. Stochastic trajectory prediction via motion indeterminacy diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17113–17122, 2022.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Junfeng Hu, Xu Liu, Zhencheng Fan, Yuxuan Liang, and Roger Zimmermann. Towards unifying diffusion models for probabilistic spatio-temporal graph learning. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, pp. 135–146, 2024.

John L Keltner, Chris A Johnson, Kimberly E Cello, Mary A Edwards, Shannan E Bandermann, Michael A Kass, Mae O Gordon, Ocular Hypertension Treatment Study Group, et al. Classification of visual field abnormalities in the ocular hypertension treatment study. *Archives of Ophthalmology*, 121(5):643–650, 2003.

Ruilong Li, Shan Yang, David A Ross, and Angjoo Kanazawa. Ai choreographer: Music conditioned 3d dance generation with aist++. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 13401–13412, 2021.

Yan Li, Xinjiang Lu, Yaqing Wang, and Dejing Dou. Generative time series forecasting with diffusion, denoise, and disentanglement. *Advances in Neural Information Processing Systems*, 35: 23009–23022, 2022.

Kai Lv, Liang Yuan, and Xiaoyu Ni. Learning autoencoder diffusion models of pedestrian group relationships for multimodal trajectory prediction. *IEEE Transactions on Instrumentation and Measurement*, 2024.

Kedi Lyu, Haipeng Chen, Zhenguang Liu, Beiqi Zhang, and Ruili Wang. 3d human motion prediction: A survey. *Neurocomputing*, 489:345–365, 2022.

Giovanni Montesano, Andrew Chen, Randy Lu, Cecilia S Lee, and Aaron Y Lee. Uwhvf: a real-world, open source dataset of perimetry tests from the humphrey field analyzer at the university of washington. *Translational Vision Science & Technology*, 11(1):2–2, 2022.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International conference on machine learning*, pp. 8857–8868. PMLR, 2021.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems*, 34:24804–24816, 2021.

Ye Tian, Mingyang Zang, Anurag Sharma, Sophie Z Gu, Ari Leshno, and Kaveri A Thakoor. Glaucoma progression detection and humphrey visual field prediction using discriminative and generative vision transformers. In *International Workshop on Ophthalmic Medical Image Analysis*, pp. 62–71. Springer, 2023.

Weizhuo Wang, C Karen Liu, and Monroe Kennedy III. Egonav: Egocentric scene-aware human trajectory prediction. *arXiv preprint arXiv:2403.19026*, 2024.

Haomin Wen, Youfang Lin, Yutong Xia, Huaiyu Wan, Qingsong Wen, Roger Zimmermann, and Yuxuan Liang. Diffstg: Probabilistic spatio-temporal graph forecasting with denoising diffusion models. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, pp. 1–12, 2023.

Theodor Westny, Björn Olofsson, and Erik Frisk. Diffusion-based environment-aware trajectory prediction. *arXiv preprint arXiv:2403.11643*, 2024.

Xiulong Yang, Sheng-Min Shih, Yinlin Fu, Xiaoting Zhao, and Shihao Ji. Your vit is secretly a hybrid discriminative-generative diffusion model. *arXiv preprint arXiv:2208.07791*, 2022.

Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, et al. A survey on diffusion models for time series and spatio-temporal data. *arXiv preprint arXiv:2404.18886*, 2024.

Tianlong Zhao, Guangle Song, Xuemei Li, Lizhen Cui, and Caiming Zhang. Diff-mgr: Dynamic causal graph attention and pattern reproduction guided diffusion model for multivariate time series probabilistic forecasting. *Information Sciences*, 675:120742, 2024.

A Proof of Theorem 1

For convenience, let $X = x_{1:T}$ and $Y = x_{T:T'}$. Then we have

$$\begin{aligned} L_{f_G}(X) &= \|Y - Af_A\left(c_A(X)\right)\| \\ &= \|Y - Af_A\left(c_A(X)\right) + Ac_A(Y) - Ac_A(Y)\| \\ &\geq \|Y - Ac_A(Y)\| - \|Af_A\left(c_A(X)\right) - Ac_A(Y)\| \\ &= L_{c_A}(Y) - L_{f_A}(X). \end{aligned}$$

The inequality above follows from the reverse triangle inequality.

B Proof of Theorem 2

First, observe that $u_A(x_t)$ is the distance between x_t and the set $\mathrm{Conv}A$. This can be seen by noting that

$$c_A(x_t) = \arg\min_c \|x_t - Ac\| = \arg\min_{\bar{x} \in \mathrm{Conv}A} \|x_t - \bar{x}\|, \tag{9}$$

where $c \in \mathbb{R}^p$ has positive elements that sum to one. Let $\hat{x} = Ac_A(x_t)$. Recall from Definition 1 that $u_A(x_t) = \|x_t - \hat{x}\|$.

Similarly, $\mathrm{dist}(x_t, \mathrm{Conv}D)$ is defined as

$$\mathrm{dist}(x_t, \mathrm{Conv}D) = \min_{\bar{d} \in \mathrm{Conv}D} \left\|x_t - \bar{d}\right\|.$$

Let d be such that $\mathrm{dist}(x_t, \mathrm{Conv}D) = \|x_t - d\|$.

The remainder of the proof follows from a straightforward application of the reverse triangle inequality:

$$\begin{aligned} \|\hat{x} - d\| &= \|\hat{x} - d + x_t - x_t\| \\ &\geq \left|\|x_t - d\| - \|x_t - \hat{x}\|\right|. \end{aligned}$$

Then, we have

$$-\|\hat{x}-d\| \leq \|x_t-d\| - \|x_t-\hat{x}\| \leq \|\hat{x}-d\|\,.$$

With some rearranging, we arrive at Equation equation 4 by letting $\delta = \|\hat{x}-d\|$.

Now note that if $p = n$, then selecting $A = D$ minimizes the archetypal analysis objective equation 1 (Cutler & Breiman, 1994, Proposition 1) with RSS of 0, and A fully expresses D. Then $\mathrm{Conv}A = \mathrm{Conv}D$ and $\hat{x} = d$. Finally, $\delta = 0$. We briefly remark that δ therefore captures the expressiveness of the archetypes.

C Pattern Mixing

In Lines 9 and 10 of Algorithm 2, we include an additional pattern mixing step in our sequence prediction process. The final PGDM prediction is a linear combination of the raw pattern prediction from the guidance function and the the pattern-guided output of the diffusion model. We include this step to overcome some of the practical challenges of PGDM. In practice, we find that PGDM's capacity for pattern guidance is highly dependent on appropriate architecture design. We therefore include pattern mixing as an additional step to overcome this challenge.

While pattern mixing improves prediction quality, pattern guidance is still necessary. Figure 5 illustrates the impact of pattern guidance and pattern mixing. Without guidance, the model may make highly varied predictions that are far from the groundtruth. With pattern guidance, PGDM narrows the distribution of predictions and shifts it towards the ground truth. Pattern mixing further shifts the distribution, without affecting sample diversity. Our results demonstrate exactly this. The unguided PGDM prediction has higher error and variance. With guidance and mixing, the error and standard deviation are significantly reduced, demonstrating that both pattern guidance and pattern mixing aid in improving predictions.

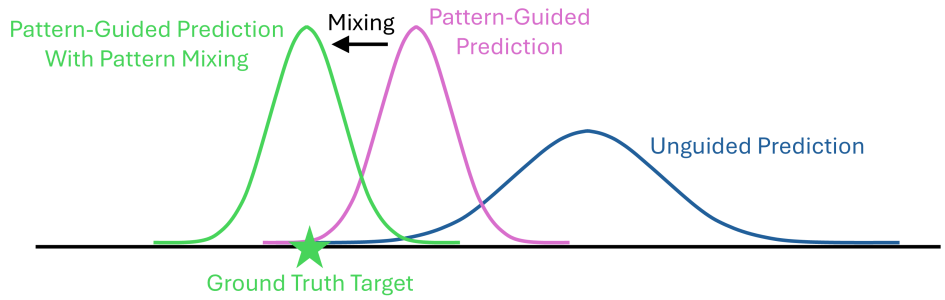


Figure 5: Impacts of pattern guidance and pattern mixing.

D Implementation Details

For both case studies, we split our data into 70% training, 15% validation, and 15% test sets. For the motion capture data, we remove rotations around the vertical axis and supply the isolated rotation angles as additional inputs to PGDM and the baseline models. This normalizes the direction in which the motion capture skeletons are facing, allowing for more straightforward pattern extraction.

To train our guidance function, we first extract archetypal patterns from the most recent VF x_T of each sequence in the training set. By extracting archetypes from only a single point in each sequence, we avoid leakage between the training, validation, and test sets. We select the number of archetypes p by a hyperparameter search through $p = 2, \ldots, 25$ with selection criterion following Elze et al. (2015). We then train our pattern prediction model (see Figure 6 for architecture) to predict the pattern representation of each sequence. We train the model with the Adam optimizer on a KL-divergence loss function with the hyperparameters shown in Table 3 and patience 20. These hyperparameters were selected over a search of batch size 32 to 64 and learning rate 10^{-4} to 5×10^{-4}, with mean absolute error (MAE) as selection criterion.

Table 3: Hyperparameter choices for pattern prediction model.

	Pattern Prediction Model		PGDM$_{\text{MAE}}$			PGDM$_{\text{GDE}}$		
	Batch Size	LR	Batch Size	LR	Epochs	Batch Size	LR	Epochs
UWHVF	32	1×10^{-4}	32	5×10^{-5}	200	64	$1 \times 10^{-}5$	100
Break	32	5×10^{-4}	32	1×10^{-3}	300	64	5×10^{-4}	200
House	64	5×10^{-4}	32	1×10^{-3}	300	32	5×10^{-4}	200
Ballet jazz	64	5×10^{-4}	32	1×10^{-3}	300	64	5×10^{-4}	300
Street jazz	64	5×10^{-4}	32	1×10^{-3}	300	64	1×10^{-3}	200
Krump	64	5×10^{-4}	32	1×10^{-3}	300	32	5×10^{-4}	200
LA Hip Hop	64	5×10^{-4}	32	1×10^{-3}	300	32	5×10^{-4}	200
Lock	64	5×10^{-4}	32	1×10^{-3}	300	32	1×10^{-3}	200
Middle Hip Hop	64	5×10^{-4}	32	1×10^{-3}	300	64	5×10^{-4}	200
Pop	32	5×10^{-4}	32	1×10^{-3}	300	32	5×10^{-4}	200
Wack	32	5×10^{-4}	32	1×10^{-3}	300	32	5×10^{-4}	300

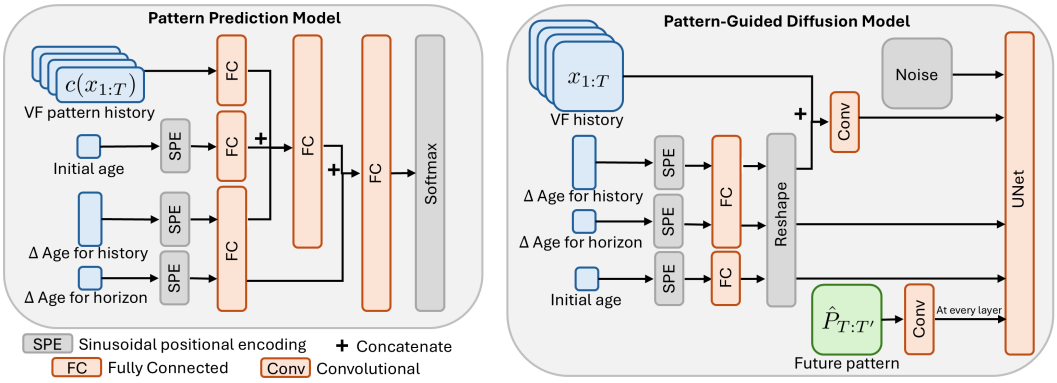


Figure 6: Pattern prediction model and pattern-guided diffusion model for visual field prediction.

We train our diffusion model (see Figure 6 for architecture) with the Adam optimizer on a mean square error loss function with the hyperparameters shown in Table 3. For the VF prediction application, these hyperparameters were selected over a search of batch size 32 to 64, learning rate 10^{-5} to 5×10^{-5}, and 100 to 1000 epochs. For the motion prediction application, the model was trained with learning rate scheduling, and the hyperparameters were selected over a search of batch size 32 to 64, learning rate 5×10^{-4} to 10^{-3} and 200 to 300 epochs. In both applications, for our selection criterion, we measure MAE with maximum guidance scale $\bar{w} = 1, \ldots, 10$ and no pattern mixing, and we choose only from models with the highest capacity for pattern guidance (i.e., error continues to reduce with increasing $\bar{w}$). Of these, we select the models with lowest achievable MAE over the tested range of $\bar{w}$. To better evaluate the full effect of pattern guidance on model performance, we also select models with the highest impact of pattern guidance over the tested range of $\bar{w}$ (e.g., the greatest achievable percent decrease in error from applying guidance). For all PGDM models, we train with conditioning dropout probability $p_{\text{drop}} = 0.2$. We evaluate with maximum tolerable uncertainty γ = 0.1, 0.03, 0.06, 0.04, 0.04, 0.05, 0.05, 0.06, 0.06, 0.03, and 0.05 for UWHVF, break, house, ballet jazz, street jazz, krump, LA hip hop, lock, middle hip hop, pop, and wack, respectively. These were chosen based on the range of uncertainties on the validation data.

For our baselines TimeGrad and CSDI, we select hyperparameters following the published implementation details. For the GenViT model, we select hyperparameters from a hyperparameter search, as those published in Tian et al. (2023) were for a simpler task with $H = 1$ and $T = 1$. We train with batch size 16, learning rate 10^{-5}, and 50 epochs. These hyperparameters were selected from a search over batch size 8 to 16, learning rate 10^{-5} to 5×10^{-5}, and 10 to 50 epochs, with MAE as selection criterion. We chose these ranges based on the hyperparameters used in Tian et al. (2023).

All experiments were performed on a machine with 42 GB of GPU memory. Each model requires less than 1 GB.

E PATTERNS EXTRACTED FROM AIST++

Figures 7 to 15 show the patterns extracted from each genre of the AIST++ dataset.

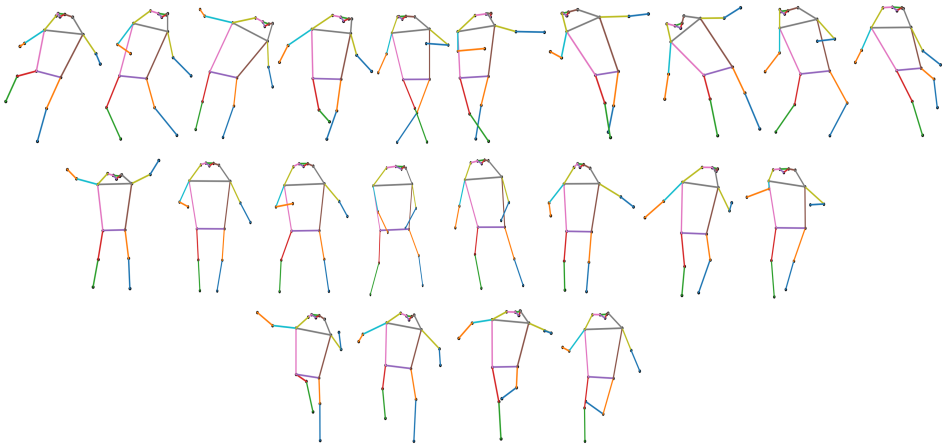


Figure 7: Twenty two archetypes extracted from AIST++ house dancing frames.

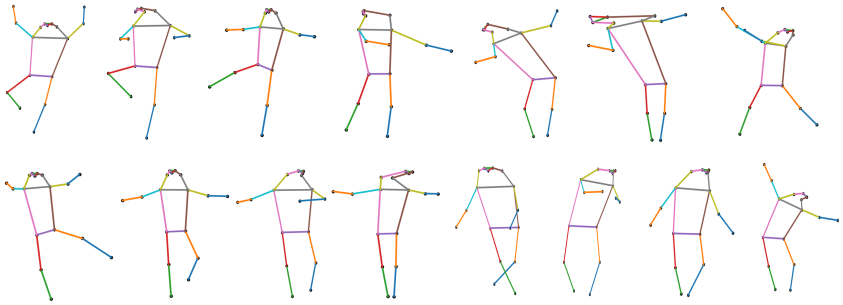


Figure 8: Fifteen archetypes extracted from AIST++ ballet jazz dancing frames.

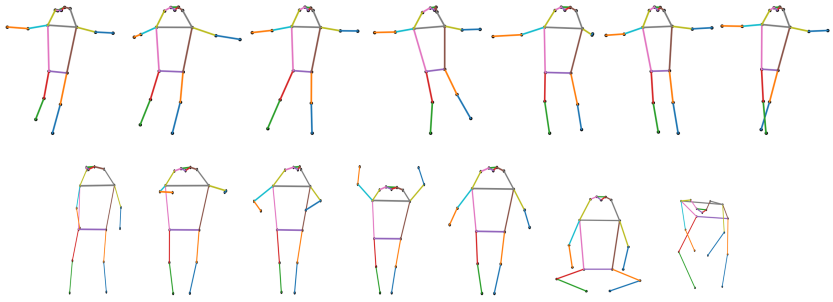


Figure 9: Fourteen archetypes extracted from AIST++ street jazz dancing frames.

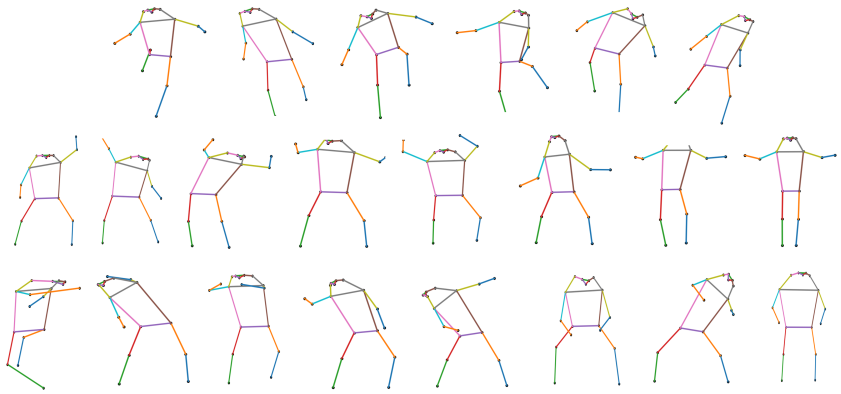


Figure 10: Twenty two archetypes extracted from AIST++ krump dancing frames.

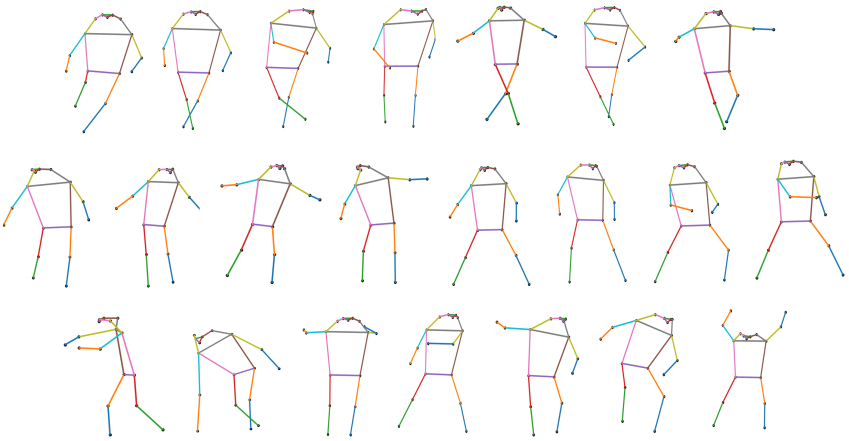


Figure 11: Twenty two archetypes extracted from AIST++ LA hip hop dancing frames.

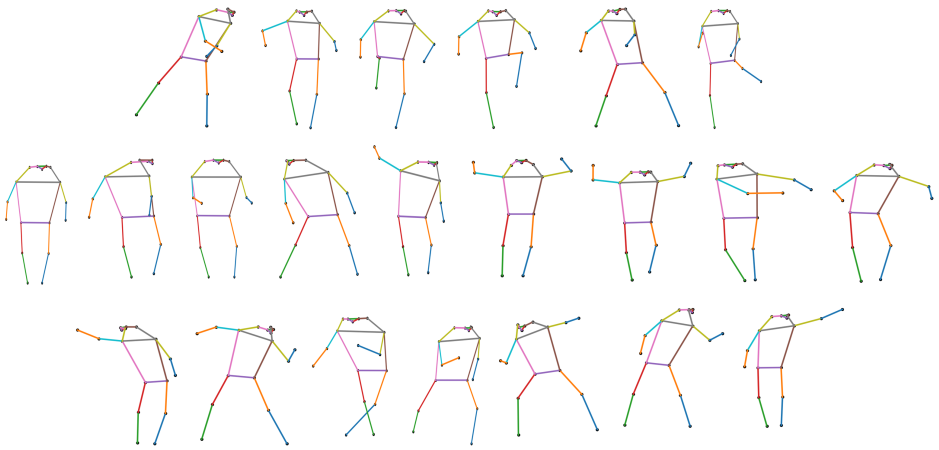


Figure 12: Twenty two archetypes extracted from AIST++ lock dancing frames.

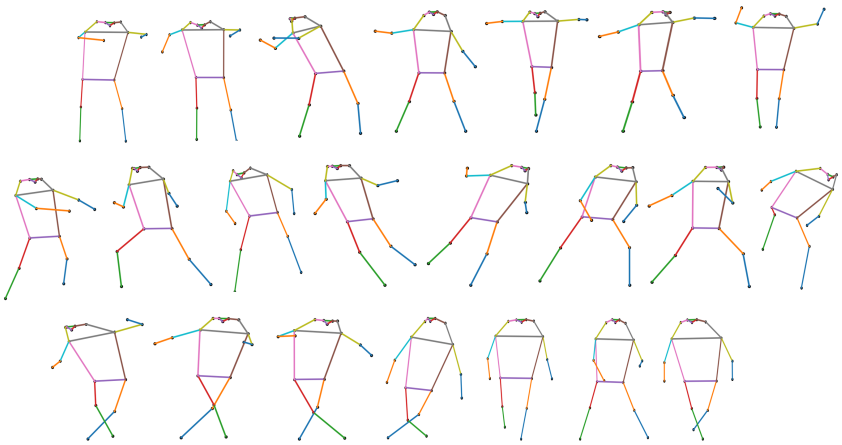


Figure 13: Twenty two archetypes extracted from AIST++ middle hip hop dancing frames.

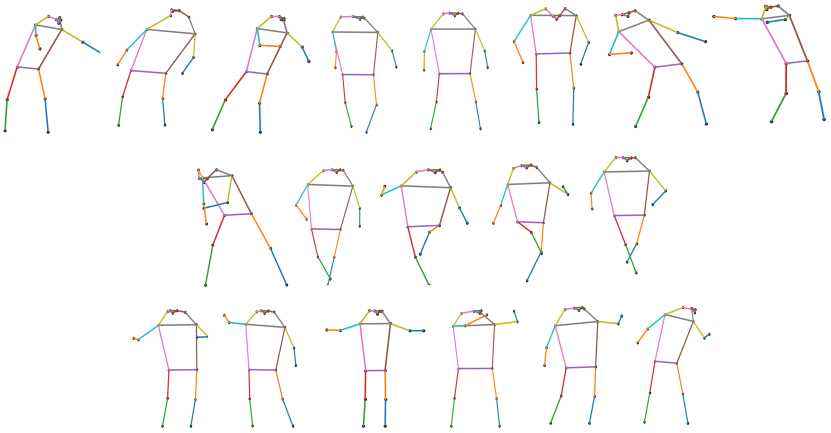


Figure 14: Nineteen archetypes extracted from AIST++ pop dancing frames.

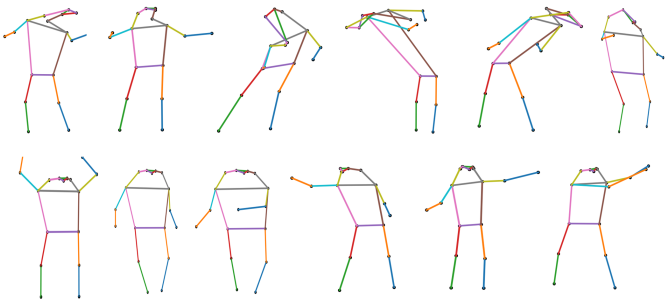


Figure 15: Twelve archetypes extracted from AIST++ wack dancing frames.

F Evaluation of PGDM Components

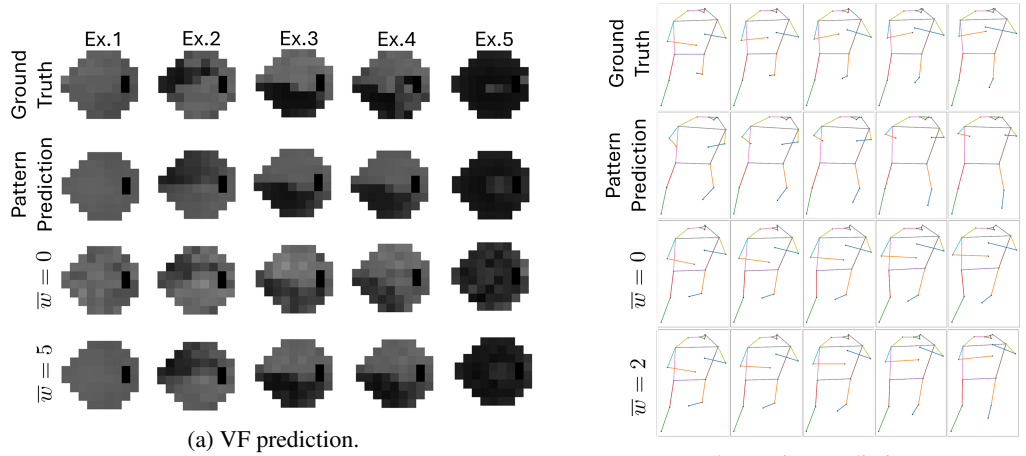
Table 4 shows the reconstruction error of the extracted patterns, the error of the pattern prediction model, and the error of the guidance function.

Table 4: Mean absolute error (MAE) of the guidance function and its components. Note that the pattern prediction is performed in the pattern representation space, with range $(0, 1)$.

	Archetypal Analysis	**Pattern Prediction**	**Guidance Function**
UWHVF	2.3146	0.0466	3.1560
break	0.7929	0.0111	0.8440
house	1.6591	0.0111	1.7636
ballet jazz	0.6269	0.0156	0.6698
street jazz	1.3009	0.0108	1.3438
krump	1.6768	0.0085	1.7677
la hip hop	1.6707	0.0107	1.7622
lock	1.3892	0.0102	1.4743
middle hip hop	1.8489	0.0113	1.9682
pop	0.9886	0.0106	1.0572
wack	0.5353	0.0166	0.5973

G Impact of Pattern Guidance

In the main text, we observed that pattern guidance reduces the error of PGDM's predictions. To further illustrate this point, qualitative examples for both applications are shown in Figure 16. For VF prediction, we show five example $H = 1$ step-ahead predictions from PGDM$_{\text{GDE}}$. When pattern guidance is not used ($\overline{w} = 0$), PGDM makes a noisy prediction based only on the past visual field data. When pattern guidance is added ($\overline{w} = 5$), PGDM incorporates the pattern prediction in its forecast. The outcome resembles a mixture of the pattern prediction and the unguided prediction (see Ex. 2 of 16a). For motion prediction, we show one example $H = 5$ step ahead prediction for PGDM$_{\text{GDE}}$. In this example, we highlight the bent right leg of the skeleton. Without pattern guidance ($\overline{w} = 0$), the model predicts nearly no motion in the leg across the horizon. In contrast, the guidance function predicts a set of patterns that change over time, matching the moving right leg of the ground truth frames. When guidance is used ($\overline{w} = 2$), PGDM incorporates this motion into its prediction and forecasts more accurate future frames.



(a) VF prediction.

(b) Motion prediction.

Figure 16: Qualitative examples of pattern guidance for PGDM$_{\text{GDE}}$ on a) the visual field prediction application and b) the human motion prediction application.

In Tables 5 to 14, we show the quantitative effect of pattern guidance levels $\overline{w} = 1, 2, 3, 4, 5$ on PGDM for the human motion prediction application. Each table corresponds to one of the dance genres in the AIST++ dataset. In general PGDM_{MAE} and PGDM_{GDE} achieve their best performances with relatively light guidance. Beyond this point, the pattern guidance has diminishing returns, even increasing the prediction error when the guidance scale is too high. In practice, the appropriate $\overline{w}$ may be selected in a manner similar to a hyperparameter search. We also observe that, in most cases, the standard deviation of the MAE decreases as $\overline{w}$ increases up to the optimal $\bar{w}$. This indicates that pattern guidance improves both the quality and the consistency of predictions.

Table 5: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the break dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		2.10±0.0064	-	-	-
CSDI		0.47±0.0010	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.41±0.0032	80.33±0.10	12.54±0.15	-
	$\overline{w} = 1$	0.39±0.0011	81.56±0.08	17.99±0.15	6.24±0.12
	$\overline{w} = 2$	0.39±0.0011	81.63±0.08	18.32±0.16	6.61±0.13
	$\overline{w} = 3$	0.40±0.0009	80.99±0.06	15.45±0.16	3.33±0.25
	$\overline{w} = 4$	0.42±0.0010	79.89±0.09	10.54±0.17	-2.28±0.31
	$\overline{w} = 5$	0.45±0.0011	78.51±0.09	4.42±0.15	-9.28±0.31
PGDM_{GDE}	$\overline{w} = 0$	0.52±0.0032	75.25±0.18	-10.06±0.76	-
	$\overline{w} = 1$	0.47±0.0022	77.61±0.13	0.41±0.62	9.51±0.31
	$\overline{w} = 2$	0.46±0.0013	78.00±0.11	2.16±0.42	11.10±0.38
	$\overline{w} = 3$	0.48±0.0006	77.26±0.09	-1.12±0.23	8.12±0.49
	$\overline{w} = 4$	0.50±0.0011	76.06±0.12	-6.48±0.19	3.25±0.68
	$\overline{w} = 5$	0.53±0.0015	74.62±0.13	-12.90±0.23	-2.58±0.81

Table 6: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the house dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		3.71±0.0159	-	-	-
CSDI		1.02±0.0045	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.79±0.0017	78.63±0.12	22.11±0.35	-
	$\overline{w} = 1$	0.74±0.0012	80.04±0.11	27.24±0.31	6.59±0.10
	$\overline{w} = 2$	0.74±0.0011	80.12±0.11	27.53±0.29	6.96±0.09
	$\overline{w} = 3$	0.76±0.0012	79.47±0.12	25.16±0.28	3.91±0.17
	$\overline{w} = 4$	0.80±0.0014	78.39±0.12	21.25±0.34	-1.11±0.20
	$\overline{w} = 5$	0.85±0.0019	77.05±0.14	16.35±0.42	-7.40±0.29
PGDM_{GDE}	$\overline{w} = 0$	0.90±0.0013	75.64±0.11	11.19±0.39	-
	$\overline{w} = 1$	0.83±0.0012	77.60±0.12	18.35±0.36	8.06±0.11
	$\overline{w} = 2$	0.82±0.0017	77.80±0.12	19.10±0.40	8.90±0.25
	$\overline{w} = 3$	0.85±0.0017	77.13±0.13	16.62±0.42	6.11±0.23
	$\overline{w} = 4$	0.89±0.0019	76.03±0.14	12.64±0.29	1.63±0.17
	$\overline{w} = 5$	0.94±0.0014	74.75±0.13	7.96±0.29	-3.64±0.14

Table 7: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the ballet jazz dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		1.32±0.0098	-	-	-
CSDI		0.55±0.0054	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.42±0.0008	67.94±0.19	22.83±0.84	-
	$\overline{w} = 1$	0.39±0.0002	70.00±0.22	27.79±0.71	6.43±0.14
	$\overline{w} = 2$	0.40±0.0005	69.36±0.24	26.25±0.68	4.43±0.24
	$\overline{w} = 3$	0.44±0.0006	66.91±0.28	20.34±0.75	-3.23±0.31
	$\overline{w} = 4$	0.48±0.0012	63.51±0.32	12.17±0.89	-13.82±0.46
	$\overline{w} = 5$	0.53±0.0013	59.71±0.33	3.01±1.02	-25.68±0.50
PGDM_{GDE}	$\overline{w} = 0$	0.49±0.0004	62.75±0.30	10.34±0.83	-
	$\overline{w} = 1$	0.45±0.0010	66.10±0.28	18.39±0.85	8.98±0.19
	$\overline{w} = 2$	0.45±0.0010	65.61±0.30	17.21±0.86	7.67±0.22
	$\overline{w} = 3$	0.49±0.0011	62.84±0.33	10.54±0.95	0.23±0.26
	$\overline{w} = 4$	0.54±0.0009	58.90±0.35	1.05±1.01	-10.35±0.23
	$\overline{w} = 5$	0.60±0.0012	54.27±0.41	-10.09±1.13	-22.78±0.28

Table 8: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the street jazz dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		1.65±0.0102	-	-	-
CSDI		0.56±0.0054	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.52±0.0017	68.58±0.25	6.58±0.79	-
	$\overline{w} = 1$	0.48±0.0010	70.66±0.23	12.76±0.76	6.62±0.19
	$\overline{w} = 2$	0.48±0.0008	70.87±0.22	13.39±0.80	7.29±0.18
	$\overline{w} = 3$	0.49±0.0009	70.23±0.22	11.49±0.85	5.26±0.20
	$\overline{w} = 4$	0.51±0.0009	69.07±0.23	8.03±0.91	1.56±0.22
	$\overline{w} = 5$	0.54±0.0009	67.53±0.25	3.45±0.95	-3.34±0.25
PGDM_{GDE}	$\overline{w} = 0$	0.60±0.0016	63.85±0.26	-7.49±1.02	-
	$\overline{w} = 1$	0.55±0.0007	66.66±0.23	0.88±0.99	7.79±0.19
	$\overline{w} = 2$	0.54±0.0005	67.12±0.20	2.25±0.97	9.06±0.19
	$\overline{w} = 3$	0.55±0.0005	66.41±0.21	0.12±0.96	7.08±0.17
	$\overline{w} = 4$	0.58±0.0004	65.08±0.22	-3.82±0.97	3.41±0.21
	$\overline{w} = 5$	0.60±0.0005	63.39±0.22	-8.84±1.00	-1.25±0.21

Table 9: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the krump dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		2.37±0.0067	-	-	-
CSDI		0.77±0.0017	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.77±0.0016	67.56±0.13	-0.04±0.31	-
	$\overline{w} = 1$	0.71±0.0014	70.27±0.13	8.30±0.31	8.33±0.08
	$\overline{w} = 2$	0.70±0.0013	70.40±0.13	8.70±0.31	8.73±0.12
	$\overline{w} = 3$	0.73±0.0011	69.31±0.13	5.33±0.30	5.37±0.12
	$\overline{w} = 4$	0.77±0.0010	67.53±0.13	-0.14±0.28	-0.11±0.13
	$\overline{w} = 5$	0.82±0.0008	65.26±0.12	-7.15±0.25	-7.11±0.16
PGDM_{GDE}	$\overline{w} = 0$	0.88±0.0016	62.85±0.12	-14.59±0.26	-
	$\overline{w} = 1$	0.79±0.0010	66.50±0.11	-3.32±0.24	9.84±0.07
	$\overline{w} = 2$	0.79±0.0011	66.87±0.12	-2.18±0.28	10.83±0.10
	$\overline{w} = 3$	0.81±0.0012	65.91±0.12	-5.14±0.31	8.25±0.11
	$\overline{w} = 4$	0.84±0.0013	64.39±0.13	-9.84±0.33	4.14±0.11
	$\overline{w} = 5$	0.89±0.0012	62.59±0.13	-15.39±0.34	-0.70±0.12

Table 10: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the LA hip hop dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		3.30±0.0157	-	-	-
CSDI		0.78±0.0023	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.80±0.0010	75.83±0.13	-1.95±0.36	-
	$\overline{w} = 1$	0.74±0.0007	77.55±0.11	5.31±0.33	7.12±0.05
	$\overline{w} = 2$	0.74±0.0006	77.64±0.11	5.70±0.33	7.50±0.09
	$\overline{w} = 3$	0.76±0.0007	76.90±0.11	2.59±0.35	4.45±0.11
	$\overline{w} = 4$	0.80±0.0007	75.70±0.11	-2.49±0.37	-0.53±0.12
	$\overline{w} = 5$	0.85±0.0007	74.19±0.12	-8.86±0.38	-6.78±0.13
PGDM_{GDE}	$\overline{w} = 0$	0.90±0.0009	72.82±0.15	-14.62±0.34	-
	$\overline{w} = 1$	0.82±0.0009	75.04±0.14	-5.29±0.32	8.14±0.02
	$\overline{w} = 2$	0.82±0.0005	75.16±0.12	-4.77±0.32	8.59±0.05
	$\overline{w} = 3$	0.85±0.0004	74.24±0.12	-8.66±0.33	5.19±0.08
	$\overline{w} = 4$	0.90±0.0006	72.77±0.13	-14.84±0.40	-0.19±0.12
	$\overline{w} = 5$	0.96±0.0007	71.00±0.15	-22.33±0.45	-6.73±0.12

Table 11: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the lock dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		3.03±0.0086	-	-	-
CSDI		0.76±0.0028	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.72±0.0011	76.12±0.10	4.24±0.23	-
	$\overline{w} = 1$	0.67±0.0005	78.07±0.06	12.08±0.29	8.18±0.12
	$\overline{w} = 2$	0.67±0.0003	77.95±0.06	11.60±0.32	7.68±0.16
	$\overline{w} = 3$	0.70±0.0007	76.89±0.07	7.35±0.29	3.25±0.14
	$\overline{w} = 4$	0.75±0.0009	75.31±0.09	0.99±0.27	-3.40±0.13
	$\overline{w} = 5$	0.81±0.0012	73.34±0.09	-6.90±0.37	-11.63±0.25
PGDM_{GDE}	$\overline{w} = 0$	0.78±0.0017	74.17±0.09	-3.57±0.29	-
	$\overline{w} = 1$	0.71±0.0022	76.51±0.12	5.80±0.23	9.04±0.22
	$\overline{w} = 2$	0.71±0.0021	76.52±0.11	5.86±0.35	9.11±0.31
	$\overline{w} = 3$	0.75±0.0015	75.44±0.11	1.53±0.24	4.92±0.19
	$\overline{w} = 4$	0.80±0.0013	73.71±0.10	-5.40±0.26	-1.76±0.13
	$\overline{w} = 5$	0.87±0.0008	71.47±0.09	-14.39±0.34	-10.45±0.17

Table 12: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the middle hip hop dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		3.35±0.0113	-	-	-
CSDI		1.04±0.0048	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.88±0.0014	73.79±0.11	15.40±0.46	-
	$\overline{w} = 1$	0.82±0.0009	75.52±0.10	20.99±0.44	6.61±0.09
	$\overline{w} = 2$	0.82±0.0015	75.52±0.09	20.99±0.43	6.60±0.10
	$\overline{w} = 3$	0.85±0.0004	74.59±0.09	17.99±0.42	3.06±0.17
	$\overline{w} = 4$	0.90±0.0007	73.10±0.11	13.18±0.46	-2.63±0.18
	$\overline{w} = 5$	0.96±0.0015	71.21±0.13	7.10±0.48	-9.82±0.18
PGDM_{GDE}	$\overline{w} = 0$	1.05±0.0034	68.67±0.13	-1.11±0.37	-
	$\overline{w} = 1$	0.96±0.0040	71.31±0.12	7.40±0.26	8.42±0.22
	$\overline{w} = 2$	0.96±0.0033	71.49±0.06	7.99±0.33	9.00±0.31
	$\overline{w} = 3$	0.99±0.0020	70.51±0.05	4.83±0.30	5.87±0.31
	$\overline{w} = 4$	1.04±0.0030	68.92±0.05	-0.30±0.29	0.80±0.39
	$\overline{w} = 5$	1.10±0.0019	67.03±0.08	-6.40±0.37	-5.23±0.21

Table 13: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the pop dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		2.55±0.0105	-	-	-
CSDI		0.70±0.0053	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.47±0.0008	81.57±0.09	32.95±0.58	-
	$\overline{w} = 1$	0.44±0.0017	82.54±0.12	36.49±0.72	5.27±0.25
	$\overline{w} = 2$	0.44±0.0015	82.54±0.11	36.47±0.64	5.24±0.17
	$\overline{w} = 3$	0.46±0.0013	81.82±0.09	33.88±0.54	1.38±0.20
	$\overline{w} = 4$	0.49±0.0018	80.60±0.12	29.43±0.65	-5.26±0.23
	$\overline{w} = 5$	0.54±0.0019	78.99±0.12	23.59±0.73	-13.97±0.23
PGDM_{GDE}	$\overline{w} = 0$	0.52±0.0013	79.44±0.09	25.21±0.57	-
	$\overline{w} = 1$	0.49±0.0015	80.70±0.07	29.81±0.36	6.16±0.23
	$\overline{w} = 2$	0.49±0.0007	80.83±0.08	30.25±0.49	6.75±0.17
	$\overline{w} = 3$	0.50±0.0010	80.19±0.08	27.93±0.57	3.65±0.12
	$\overline{w} = 4$	0.54±0.0017	78.99±0.09	23.56±0.56	-2.20±0.31
	$\overline{w} = 5$	0.58±0.0015	77.39±0.11	17.74±0.59	-9.98±0.35

Table 14: Mean absolute error (MAE) of PGDM_{MAE}, PGDM_{GDE}, and baselines on the wack dancing genre of the AIST++ dataset. Percent improvements over baselines are shown in the Δ MAE (%) columns. Mean and standard deviation are taken across five samples.

Model		**MAE (dB)**	**Δ MAE (%) vs. TimeGrad**	**Δ MAE (%) vs. CSDI**	**Δ MAE (%) vs. $\overline{w} = 0$**
TimeGrad		1.03±0.0047	-	-	-
CSDI		0.44±0.0042	-	-	-
PGDM_{MAE}	$\overline{w} = 0$	0.44±0.0044	57.55±0.45	0.61±1.39	-
	$\overline{w} = 1$	0.41±0.0031	59.74±0.32	5.73±1.47	5.15±0.66
	$\overline{w} = 2$	0.42±0.0028	58.75±0.29	3.41±1.45	2.82±0.79
	$\overline{w} = 3$	0.45±0.0019	56.21±0.23	-2.52±1.26	-3.16±0.73
	$\overline{w} = 4$	0.48±0.0015	52.93±0.20	-10.20±1.13	-10.88±0.82
	$\overline{w} = 5$	0.52±0.0020	49.26±0.27	-18.78±1.01	-19.53±0.91
PGDM_{GDE}	$\overline{w} = 0$	0.49±0.0054	52.11±0.44	-12.12±1.81	-
	$\overline{w} = 1$	0.47±0.0083	54.15±0.67	-7.36±2.59	4.26±0.79
	$\overline{w} = 2$	0.48±0.0082	52.83±0.64	-10.45±2.48	1.49±0.83
	$\overline{w} = 3$	0.52±0.0077	49.71±0.58	-17.76±2.43	-5.02±0.64
	$\overline{w} = 4$	0.56±0.0060	45.86±0.46	-26.77±2.18	-13.06±0.30
	$\overline{w} = 5$	0.60±0.0045	41.48±0.28	-37.02±1.90	-22.21±0.64