DenseKoopman: A Plug-and-Play Framework for Dense Pedestrian Trajectory Prediction

Xianbang Li\textsuperscript{123}, Yilong Ren\textsuperscript{124*}, Han Jiang\textsuperscript{123}, Haiyang Yu\textsuperscript{123}, Yanlei Cui\textsuperscript{23} and Liang Xu\textsuperscript{23*}

\textsuperscript{1}School of Transportation Science and Engineering, Beihang University
\textsuperscript{2}State Key Lab of Intelligent Transportation System, Beihang University
\textsuperscript{3}Hangzhou Innovation Institute, Beihang University

\{1450552846, yilongren, buaajh, hyyu\}@buaa.edu.cn, \{yanleicui, xuliangccea\}@foxmail.com

Abstract

Pedestrian trajectory prediction has emerged as a core component of human–robot interaction and autonomous driving. Fast and accurate prediction of surrounding pedestrians contributes to making decisions and improves safety and efficiency. However, pedestrians’ future trajectories will interact with their surrounding traffic participants. As the density of pedestrians increases, the complexity of such interactions also increases significantly, leading to an inevitable decrease in the accuracy of pedestrian trajectory prediction. To address this issue, we propose DenseKoopman, a plug-and-play framework for dense pedestrian trajectory prediction. Specifically, we introduce the Koopman operator theory to find an embedding space for a global linear approximation of a nonlinear pedestrian motion system. By encoding historical trajectories as linear state embeddings in the Koopman space, we transform nonlinear trajectory data for pedestrians in dense scenes. This linearized representation greatly reduces the complexity of dense pedestrian trajectory prediction. Extensive experiments on pedestrian trajectory prediction benchmarks demonstrate the superiority of the proposed framework. We also conducted an analysis of the data transformation to explore how our DenseKoopman framework works with each validation method and uncovers motion patterns that may be hidden within the trajectory data. Code is available at https://github.com/lixianbang/DenseKoopman.

1 Introduction

Pedestrian trajectory prediction aims to predict pedestrians’ future trajectories based on their past trajectories and surrounding scene information. This task is essential in a variety of applications, such as human–robot interaction [Zhang et al., 2024], autonomous driving [Wang et al., 2022], and augmented reality [Yu et al., 2023]. The future trajectories of pedestrians are characterized by uncertainty, as individuals have the ability to alter their intended motion and adjust their movement direction in response to their surroundings. In general, solving this problem requires the model to learn how pedestrians move.

Although great progress has been made in recent years [Jeon et al., 2023; Zhang et al., 2023; Mohamed et al., 2022], pedestrian trajectory prediction is still affected by many unpredictable factors, such as different pedestrian walking posture habits, nonlinear motion systems, unknown intentions, and dense scenarios. All these factors can significantly reduce the accuracy and performance of pedestrian trajectory prediction. Among these disadvantages, pedestrian scene densification is perhaps the most intractable. Previous studies have proven that accuracy drops sharply as the density within a scene increases [Shi et al., 2019; Lisotto et al., 2019].

To address these challenges, numerous deep learning models have been proposed for predicting pedestrian trajectories in crowded scenarios, and these models have demonstrated excellent performance. For instance, several methods [Helbing and Molnár, 1995; Alahi et al., 2016] exploit social force to describe the self-organization of pedestrian behavior and propose many ingenious neural networks to imitate social interactions, while other methods [Dendorfer et al., 2021; Chen et al., 2021] apply the generative model to spread the distribution over all possible future trajectories or encode the multimodal distribution of future trajectories. These methods achieve high prediction performance within sparse scenes.
For both the social force and the generative model, when the density of pedestrians in the scene increases, the dynamic interaction characteristics of pedestrians become more complex and uncertain. Therefore, the difficulty of calculating the social relationships of a large number of pedestrians in the crowd and generating multiperson and multimodality prediction trajectories increases sharply, and the calculation accuracy is greatly affected.

Inspired by works on dynamical analysis of highway traffic [Avila and Mezić, 2020], we realized that Koopman theory has the potential to process data and improve model performance for temporal features in dense scenarios because it can construct state simulation transformations of nonlinear motion systems without needing to model their dynamics. Instead of spending much computing power to directly calculate and reason about the expression of the nonlinear system. Therefore, unlike most existing deep learning methods, we explore the nonlinear system of the original space from the perspective of a high-dimensional linear system. More specifically, we introduce Koopman theory, a key theoretical and popular tool in the field of fluid mechanics. In fact, in the field of various engineering applications, there are already many tasks that introduce and apply this theory, such as wind engineering [Li et al., 2023], soft robots [Bruder et al., 2021a], and model predictive control [Zhang et al., 2022]. In the field of pedestrian trajectory prediction, no work has introduced Koopman theory to predict pedestrian trajectories.

As shown in Figure 1, the Koopman theory is centred on the methodical linear portrayal of nonlinear systems, which provides a new way of representing pedestrian trajectories in extremely crowded scenarios. We propose a plug-and-play framework for dense pedestrian trajectory prediction by approximating the Koopman operator. First, the trajectories of the pedestrians in the scene are fed into our DenseKoopman encoder for space mapping, and the linear trajectory data in each Koopman space are obtained. Then, we apply various prediction algorithms to train and predict linear data within each of the Koopman spaces. Finally, we input the prediction results obtained from each Koopman space into our DenseKoopman decoder and obtain the future trajectories of pedestrians in the original space after superposition. In summary, this work has the following three major contributions.

- We incorporate the Koopman theory for nonlinear transformation in pedestrian trajectory prediction. To the best of our knowledge, this is the first study employing Koopman mode decomposition in this context.
- We propose a plug-and-play pedestrian trajectory prediction framework based on the Koopman theory, which can extract deeper pedestrian interaction features and achieve better pedestrian trajectory prediction in extremely crowded scenarios.
- We verify the applicability and effectiveness of DenseKoopman on three prediction models, which are representative models of the CNN, LSTM, and Stable Diffusion frameworks, respectively. The experimental results on three public datasets, which contain dense pedestrian scenes, prove the effectiveness of our DenseKoopman framework.

2 Related Works

2.1 Pedestrian Trajectory Prediction

Existing methods predict future trajectories through learning past trajectory patterns. In the beginning stages of this discipline, researchers used a knowledge-based approach to consider pedestrians’ physical and social factors. For instance, the SFM [Helbing and Molnár, 1995] introduces social force to describe the self-organization of pedestrian behavior. With the rapid development of deep learning, researchers have proposed many ingenious neural networks to imitate social interactions. In [Alahi et al., 2016], Social LSTM was designed to fuse neighboring nodes and generate multiple LSTM networks to obtain information about surrounding nodes by sharing the corresponding information of each node. Following Social LSTM, Trajectron++ [Salzmann et al., 2020] incorporating agent dynamics and heterogeneous data to forecasts the trajectories of a general number of diverse agents. In addition, Social-Implicit [Mohamed et al., 2022] cluster the observed trajectories based on their maximum observed speed and then generate the final output by combining the local and global streams via self-learning weights. Moreover, MID [Gu et al., 2022] formulate the trajectory prediction task as a reverse process of motion indeterminacy diffusion that can progressively discard indeterminacy from all the walkable areas until reaching the desired trajectory.

As the human density in the prediction scene increases, the computational cost of the prediction algorithm for crowd interaction modelling will also increase significantly, leading to a notable impact on the prediction accuracy. Several recent methods have been proposed to address the issue of predicting pedestrian trajectories in extremely crowded scenarios [Lisotto et al., 2019]. However, the effectiveness of these methods is not readily apparent, and a comprehensive algorithm framework for widely applicable crowded pedestrian trajectory prediction has yet to be established.

2.2 Koopman Mode Decomposition

Koopman theory is first proposed by Koopman [Koopman and Neumann, 1932] and has been widely applied in the field of fluid mechanics [Batchelor, 1967; Mezić, 2013]. Koopman mode decomposition, as known as KMD for the sake of brevity, is a set of infinite dimensional linear operators that describe the temporal evolution of observable (measurable) quantities in system dynamics [Bagheri, 2013; Williams et al., 2015]. Recently, KMD has attracted much attention due to state-of-the-art performance in various fields, including computational chemistry [Linscott et al., 2023; Tsuneda et al., 2010; Wu et al., 2017], robotics [Abraham and Murphey, 2019; Berger et al., 2015; Bruder et al., 2021b], computer vision [Wang et al., 2023; Zhang et al., 2021; Kostic et al., 2022; Mitjans et al., 2022]. However, the application of KMD to pedestrian trajectory prediction has not been explored in any research.

Our Approach. In this paper, we first introduce KMD to the pedestrian trajectory prediction field. Additionally, our work is the first to present a new data-driven-based framework to formulate the trajectory prediction task as linear prediction in Hilbert space.
3 Proposed Method

In this section, we introduce our DenseKoopman method, which models pedestrian trajectory prediction tasks in dense crowd scenarios via Koopman mode decomposition.

Specifically, we introduce the Koopman operator theory to find an embedding space for a global linear approximation of a nonlinear pedestrian motion system. By encoding historical trajectories as linear state embeddings in the Koopman space, we transform nonlinear trajectory data for pedestrians in dense scenes. Then, we describe how the framework can be applied to upgrade general purpose algorithms to be applicable to dense crowd scenarios. Finally, we present the detailed framework architecture of our method in Figure 2.

3.1 Classic Koopman Theory

Definition 1. Before delving into the details of the proposed DenseKoopman framework, it is important to first provide a brief overview of the classic Koopman theory. Suppose that a discrete-time nonlinear dynamical system $F$ can be represented by

$$F: a_t = S_t(a_0), A \in \mathcal{M}, t \in \mathcal{R},$$

where $a_0, a_t \in A$ are state vectors on the state space $\mathcal{M}$, and $S_t$ represents the solution of the dynamical system $F$. For a specific feature representation $a_0$, the dynamics typically exhibit high-level non-linearity. The Koopman operator is a linear operator, and its spectrum enables the study of nonlinear, complex, and high-dimensional systems using linear techniques, albeit in an infinite-dimensional space.

Formally, for any function $f$ and time $t$ the action of the time-$t$ Koopman operator on $f$ can be expressed as follows [Koopman, 1931; Koopman and Neumann, 1932].

$$U^t f(a_0) = f \circ S_t(a_0) = f(a_t), f \in \mathcal{H}. \quad (2)$$

Where $\mathcal{H}$ is a Hilbert space of the function $f$, and $U^t$ is the Koopman semigroup of operators, parameterized by time $t$, which compose the functions in $\mathcal{H}$ with the solution $S_t$.

Due to the infinite dimensionality of the Koopman operator, its spectrum has the potential to include a continuous spectrum alongside the conventional point spectrum found in finite-dimensional operators. Assuming that the Koopman operator of a specific dynamical system contains only a point spectrum, its evolution in time is as follows [Mezić and Banaszuk, 2004].

$$U^t \varphi(a_0) = \varphi \circ S_t(a_0) = e^{\lambda t} \varphi(a_0). \quad (3)$$

Where $\varphi \in \mathcal{H}$ is an eigenfunction of $U^t$, $e^{\lambda t}$ is a Koopman eigenvalue for $\lambda \in \mathcal{C}$. Moreover, a subspace $\mathcal{A} \subset \mathcal{H}$ remains unchanged by the dynamics if, for any $f \in \mathcal{A}$, its image under the flow, $U^t f \in \mathcal{A}$, remains constant for all time $t$.

If the Koopman eigenfunctions, or a subset of them, constitute a basis for an invariant subspace of $\mathcal{H}$, then any function within the invariant subspace can be represented using the eigenfunction basis [Mezić and Banaszuk, 2004].

Definition 2. Formally, let $\phi = \{\varphi_i\}, i \in \mathcal{N}$ be a set of Koopman eigenfunctions and let $\mathcal{E}_\phi = \text{span}(\phi)$ be the Koopman eigenspace linked to the eigenfunction basis. The Koopman mode decomposition of any observable $F$ can be formally expressed as follows.

$$U^t f(a_0) = f \circ S_t(a_0) = \sum_{i=1}^{\infty} \varphi_i(a_0)e^{\lambda_i t} v_i. \quad (4)$$

Where $v_i = \langle f, \varphi_i^* \rangle$ is the skew projection of $f$ onto $\mathcal{E}_\phi$, obtained via the inner product of $f$ and the dual eigenfunction.
φ, and is called a Koopman mode [Mezić and Banaszuk, 2004]. The finite-dimensional approximation of the Koopman operator \( U^* \) is commonly achieved through the use of mode decomposition [Schmid, 2010].

### 3.2 Problem Formulation

The objective of pedestrian trajectory prediction is to produce credible future trajectories for pedestrians by analysing their previous movements. The prediction system takes \( N \) historical trajectories in a scene as input and generates \( N \) future trajectories in the same scene as output.

\[
x^i = \{ s^i_t \in \mathbb{R}^2 | t = -T_{past}, -T_{past} + 1, \cdots, 0 \}.
\]

\[
y^i = \{ s^i_t \in \mathbb{R}^2 | t = 1, 2, \cdots, T_{future} \}.
\]

(5)

Where the \( s^i_t \) is the 2D location at timestamp \( t \). \( T_{past} \) indicates the length of the observed trajectory, and the current timestamp is \( t = 0 \).

### 3.3 Trajectory Mode Decomposition

**Definition 3.** The following section delves into the specific application of Koopman mode decomposition to pedestrian trajectory prediction. The past trajectory of \( N \) pedestrians in time \( T_{past} \) can be expressed as a time-ordered data matrix \( X \) that contains a total of \( N \) data vectors. The 2D position coordinates of the pedestrian at an instant in time \( t \) correspond to the \( t \)-th column of the past trajectory data matrix \( X \).

\[
X = \begin{bmatrix}
L_{1,-T_{past}} & L_{1,-T_{past}+1} & \cdots & L_{1,0} \\
L_{2,-T_{past}} & L_{2,-T_{past}+1} & \cdots & L_{2,0} \\
\vdots & \vdots & \ddots & \vdots \\
L_{n,-T_{past}} & L_{n,-T_{past}+1} & \cdots & L_{n,0}
\end{bmatrix}.
\]

(6)

Where \( L \) is the 2D position coordinate of the pedestrian, with its elements forming the \( N \times T \) matrix \( X \). As previously stated, the Koopman operator \( U^* \) can be perceived as the evolution of functions within the Hilbert space of all possible \( \varphi \). However, the Koopman operator \( U^* \) is not manageable due to its infinite-dimensional nature. Up to now, numerous algorithms have been developed to approximate the eigenfunctions, eigenvalues, and modes of the Koopman operator.

**Theorem 1.** The Exact-DMD seeks to approximate a finite-dimensional representation of the Koopman operator [Schmid, 2010]. A finite-dimensional matrix \( K \) can be utilized as an approximation of the Koopman operator, the results of which can be expressed as follows.

\[
X' = KX + r.
\]

(7)

Where \( r \) is a residual error term arising from the finite-dimensional approximation of a possibly infinite expansion.

The Exact-DMD achieves this approximation by minimizing the residual term using a least squares approach. Thus, by employing the singular value decomposition (SVD) of \( X = U\Sigma W^* \) is employed to rephrase \( X' = KX + r \) as such.

\[
X' = KX + r = KU\Sigma W^* + r.
\]

(8)

By multiplying both sides of the above equation by \( U^* \) and considering that minimizing the residual term necessitates it to be orthogonal to \( U \), we derive the expression in following equation.

\[
U^*X' = U^*KU\Sigma W^* + U^*r = U^*KU\Sigma W^*.
\]

(9)

By multiplying both sides of the above equation by \( W\Sigma^{-1} \), we derive the expression in the following equation similarly.

\[
U^*X'W\Sigma^{-1} = U^*KU\Sigma W^*W\Sigma^{-1} = U^*KU \equiv S.
\]

(10)

Where \( S \) share the same eigenvalues and the eigenvectors with \( K \).

Finally, since the sampled data produced a discrete time description of an originally continuous time process, kernel mean matching (KMD) can be utilized to derive a description of the observed data points through the following equation:

\[
X_{kmd}(t) = \sum_{i=1}^{l} b_{0i} v_i e^{\omega_i t} = Ve^{\omega t}b_0
\]

(11)

Where \( l \) is the dimension of the Koopman Mode Decomposition. And \( \omega_i \) is the continuous time eigenvalue, which is given by \( \omega_i = \frac{\ln(\lambda_i)}{T} \), where \( T \) is the sampling rate and \( \lambda_i \) is the eigenvalues of \( K \) and \( S \). Here, \( V \) is a matrix whose columns are the eigenvectors \( v_i \), \( b_0 \) is a vector of coefficients associated with the initial data snapshot \( X \), and \( e^{\omega t} \) represents a diagonal matrix whose elements are \( e^{\omega_i t} \).

Consequently, it is possible to reconstruct the initial data to \( l \) dimensions within the Hilbert space. And the data in dimension \( i \) is represented as follows.

\[
X_{kmd}(i, t) = b_{0i} v_i e^{\omega_i t}
\]

(12)

**Theorem 2.** As previously stated by Koopman theory, the Koopman operator is a linear operator, and its spectrum is within an infinite-dimensional Hilbert space.

\[
X_{past}(t) = \sum_{i=1}^{l} X_{kmd}(i, t)
\]

(13)

Where \( X_{kmd}(i, t) \) is the data obtained after the decomposition of \( X_{past}(t) \). Consequently, each data point decomposed using KMD methods exhibits linearity, in contrast to the non-linearity of the original dimensions. In time series prediction tasks, linear data are generally easier to predict than nonlinear data. The above is the main role of our KMD encoder network.

### 3.4 Training and Inference Objective

After performing Koopman Mode Decomposition on the trajectory data in the original space, we obtain the corresponding finite multidimensional data. In accordance with Koopman’s theory, each dimension of the obtained data is inherently linear, thus enabling direct training of the model to learn these linear features after mode decomposition.

For nonlinear data, it is challenging for the model to learn its evolutionary characteristics and make predictions. However, for linearized data, accurate predictions are extremely
easy to obtain. Similarly, the future trajectory of \( N \) pedestrians in time \( T_{\text{future}} \) can be expressed as a time-ordered data matrix \( Y \) that contains a total of \( N \) data vectors. The 2D position coordinates of the pedestrian correspond to the future trajectory data matrix \( Y \).

\[
Y = \begin{bmatrix}
L_{1,1} & L_{1,2} & \cdots & L_{1,T_{\text{future}}} \\
L_{2,1} & L_{2,2} & \cdots & L_{2,T_{\text{future}}} \\
\vdots & \vdots & \ddots & \vdots \\
L_{n,1} & L_{n,2} & \cdots & L_{n,T_{\text{future}}}
\end{bmatrix}.
\]  
(14)

Where \( Y \) is the future trajectory, which the model need to predict. For each dimension of the decomposed data \( X_{\text{kmd}}(i, t) \), the model will produce a linear prediction \( Y_{\text{kmd}}(i, t) \).

\[
Y_{\text{kmd}}(i, t) = \text{Model}[X_{\text{kmd}}(i, t)]
\]  
(15)

Where \( \text{Model} \) is any pedestrian trajectory prediction model that can easily predict the evolution trend of linear data.

During training, we decompose the acquired data to augment the original spatial data. Each augmented datum is considered independent, and inference prediction is carried out by updating the model training parameters. Therefore, the plug-and-play framework proposed in this paper does not require changing the internal structure of the model during training, and it can be applied to any trajectory prediction model. Since the input and output of the trajectory prediction task consist of raw spatial data, a decoder is required to integrate the linear prediction data from each dimension during the reasoning and prediction process.

\[
Y_{\text{predict}}(t) = \sum_{i=1}^{l} Y_{\text{kmd}}(i, t)
\]  
(16)

Where \( Y_{\text{kmd}}(i, t) \) is the prediction of the data \( X_{\text{kmd}}(i, t) \), and \( X_{\text{kmd}}(i, t) \) is obtained after the decomposition of \( X_{\text{test}}(t) \). Additionally, the Koopman Mode Decomposition decoder network plays the main role in our work.

### 3.5 Network Architecture

Different from other trajectory prediction models that directly train and predict nonlinear pedestrian motion trajectories, we design a new Koopman-based network for our DenseKoopman model. With the Koopman mode decomposition, the model can easily learn the evolutionary pattern of linearized data and make accurate predictions. Specifically, DenseKoopman consists of two key networks: a Koopman-based encoder that can linearize the trajectory data in a data-driven manner and a decoder for trajectory mode combination. An overview of the entire architecture is illustrated in Figure 2. We introduce each part in detail in the following sections.

The Koopman-based encoder maps the historical trajectory of pedestrians in a scene to an infinite-dimensional Hilbert space and uses the Koopman Mode Decomposition method for decomposition to reduce the dimensionality and obtain a limited number of corresponding linearized enhanced data \( X_{\text{kmd}}(i, t) \). These linearized enhanced data are fed into the training model as independent training data. It is important to note that designing the model for training and inferring trajectories is not the main focus of our work. DenseKoopman is a model-agnostic framework that can be directly integrated with different pedestrian trajectory prediction models introduced in previous methods.

As depicted in Figure 2, the encoder’s inputs consist of the motion trajectories of each pedestrian within the observed scene. The Koopman mode decomposition method is utilized in the encoder to process pedestrian trajectory data in the original space within the scene, resulting in the extraction of decomposed linearized data. Figure 2 clearly shows that these data exhibit clear linearization characteristics, making them amenable to learning and inference by diverse prediction models. Then, We verify the applicability and effectiveness of DenseKoopman on three prediction models, Social-Implicit [Mohamed et al., 2022], Trajectron++ [Salzmann et al., 2020], and MID [Gu et al., 2022], which are representative models of the CNN, LSTM, and Stable Diffusion frameworks, respectively. We can obtain the prediction results \( Y_{\text{kmd}}(i, t) \) of each dimension by applying these models to learn and reason about the data \( X_{\text{kmd}}(i, t) \). Finally, the prediction results \( Y_{\text{kmd}}(i, t) \) of each dimension are fed into the decoder network to obtain the future trajectory prediction of pedestrians in the original space. The decoder network is designed to integrate the linear prediction data from each dimension during the reasoning and prediction process. This superposition method can be readily applied to a variety of verification models.

### 4 Experiments

#### 4.1 Experimental Setup

**Datasets and Evaluation Metric.** Different from existing research on pedestrian trajectory prediction, we use denser datasets than ETH/UCY [Lerner et al., 2007; Pellegrini et al., 2010] to simulate extremely crowded scenarios in the real world. We evaluated our method on three public datasets containing dense pedestrian scenarios, namely, Multiple Object Tracking 20 [Dendorfer et al., 2020], Head Tracking 21 [Sundararaman et al., 2021], VSCrowd [Li et al., 2022]. We employ two widely used metrics to evaluate the prediction accuracy, the average displacement error (ADE) and the final displacement error (FDE) [Alahi et al., 2016; Pellegrini et al., 2009]. The average displacement error (ADE) was calculated as the mean discrepancy between the actual positions and the estimated positions throughout the trajectory. On the other hand, the final displacement error (FDE) measures the difference in displacement between the final positions of the actual and predicted trajectories. For each dataset, we implemented the leave-one-out cross-validation approach, where the model was trained on four scenarios and then tested on the remaining scene [Salzmann et al., 2020; Kosaraju et al., 2019]. Given the probabilistic nature of our approach, we employed a Best-of-20 strategy to calculate the ultimate ADE and FDE. The size of the video images in each dataset is [1920, 1080] pixels. In all the scenarios, the locations of pedestrians are documented in relation to the pixel coordinates, resulting in pixel measurements.
Baselines and Implementation Details. We choose three validation methods as our baseline, namely, Social-Implicit [Mohamed et al., 2022], Trajectron++ [Salzmann et al., 2020], and MID [Gu et al., 2022], which are representative of the CNN, LSTM and stable diffusion methods, respectively, for the pedestrian trajectory prediction task. The paths are sampled at intervals of 0.4 seconds, with the initial 3.2 seconds of a path serving as observed data for forecasting the subsequent 4.8 seconds of future trajectory. Therefore, we anticipate the future trajectories of pedestrians over a time frame of 12 frames based on their observed movements over a period of 8 frames. The sampling of trajectories is conducted at a frame rate of 2.5 frames per second. To verify the superiority of the proposed framework in crowded scenarios containing at least 20 people, we process data with [20 pedestrians, 8 seconds] as the spatiotemporal window.

4.2 Main Results
We apply our framework to current methods for quantitative comparison. Table 1 shows the comparison of these methods before and after applying our framework to three public datasets containing dense pedestrian scenarios. Obviously, for extremely dense pedestrian scenarios, existing pedestrian trajectory prediction algorithms involve large errors, and accurately predicting the motion trajectories of pedestrians is difficult.

Specifically, after applying our DenseKoopman framework, the original algorithm is greatly improved in terms of both the ADE and FDE accuracy metrics. This indicates that the enhanced accuracy in prediction results from improved performance in the Koopman Mode Decomposition, rather than from the use of training optimization techniques. It is worth noting that, regardless of the type of verification algorithm, the FDE significantly exceeds the ADE before applying our DenseKoopman framework. However, after applying our DenseKoopman framework, while the ADE and FDE are significantly reduced, the FDE is no longer significantly larger than the ADE.

4.3 Koopman Mode Analysis
To explore how DenseKoopman works with each validation method and uncovers motion patterns that may be hidden within the trajectory data, we analyse the data transformation before and after its passage through the encoder.

In Figure 3, we consider how the trajectory of a single pedestrian changes before and after passing through the encoder and find that the six leading patterns decomposed all have similar structures, showing a clear trend of periodic changes. As the mode decomposition progresses, the linearization becomes more pronounced. Moreover, Figure 4 plots the distribution of crowd movement velocities within the same scene. The crowd speed in the original space is chaotic, and even when the movement speed of the same person is not uniform, it shows strong irregularity and nonlinear changes. However, after passing through our KMD encoder, the velocity data exhibit similar uniform group characteristics in each of the Koopman modes. The ability to uncover such growing and decaying patterns is the significance of DenseKoopman.

4.4 Ablation Studies
In this subsection, we conducted ablation studies to investigate the effectiveness of each key component, including Zero-centered and Hankel-DMD.

Zero-centered. To examine the importance of the zero-centered approach in our framework, we reduced our zero-centered approach to the Z-score or deleted the zero-centered approach. Groups 4, 5 and 6 in Table 2 show the performance comparison. Using the same method and datasets but without our Zero-centered approach, the results degrade significantly, demonstrating the effectiveness of our Zero-centered approach. In addition, replacing the Zero-centered module in our framework with the Z-score module leads to a decrease in accuracy, as shown for Group 5 and 6 in Table 2.
Figure 3: The left side shows the spatiotemporal distribution of a single pedestrian trajectory in the original space, and its motion shows uncertainty and nonlinearity. The six graphs on the right show the spatial and temporal distributions of the data in different mode spaces after our KMD encoder, revealing an obvious linear evolution trend.

Figure 4: The left side shows the spatiotemporal distribution of all pedestrian motion velocities in the original space in the extremely crowded scene, and the crowd motion velocities are unevenly distributed with no obvious rule to say. The ten graphs on the right show the spatial and temporal distributions of crowd motion velocity data in different mode spaces after our KMD encoder, revealing a very obvious linear evolution trend and crowd characteristics.

**Hankel-DMD.** We also conducted experiments to verify the necessity of Hankel-DMD. According to Groups 2, 3, 5 and 6 in Table 2, the method with Hankel-DMD outperforms the method without Hankel-DMD by a large margin, regardless of whether a zero-centered method was used previously. This indicates that Hankel-DMD is effective at obtaining the nonlinear characteristics of trajectories from our DenseKoopman framework. Specifically, groups 1 and 4, 2 and 5, and 3 and 6 are the three types of contrast items. The only variable is whether to apply Hankel-DMD, and by comparing the performance of these three items, it is easy to determine the necessity of Hankel-DMD.

**5 Conclusions**

In this work, we propose the dense pedestrian trajectory prediction framework DenseKoopman to formulate trajectory prediction in extremely crowded scenarios. In this framework, we design a pedestrian trajectory encoder for the Koopman space transformation to obtain a linear representation of the nonlinear original trajectory data in each space. And we apply the existing prediction method to learn and infer linear transformation data in each Koopman space. Finally, we use a corresponding decoder to overlay the linear data and output the prediction of the future trajectory of the pedestrian. The experimental results demonstrate the superiority of our method, which achieves state-of-the-art performance on three public datasets within extremely crowded scenarios. DenseKoopman could be applied to a wide range of applications with autonomous driving [Jiang et al., 2024]. With linear transformation, we can obtain the linear representation of the nonlinear original trajectory, which helps the predict model to generate accurate and reasonable future trajectories. Besides, DenseKoopman is able to handle other types of spatiotemporal sequence data, which has the potential to be applied in traffic flow prediction [Ren et al., 2022], human-robots interaction and nonlinear system. In addition, it is theoretically feasible to verify other methods of mode decomposition or dimension reduction. We leave it as future work to explore the feasibility of other decomposition methods for application in dense pedestrian scenes. Despite the accurate performance and generalization, the computational cost at model training process could be expensive due to multiple Koopman data spaces. Besides, the failure of linearization may occur when the data lacks credibility and contains excessive noise.

**Acknowledgments**

This work is supported in part by Zhejiang Provincial Natural Science Foundation of China under Grant No.LD24F020008.
References


[Lisotto et al., 2019] Matteo Lisotto, Pasquale Coscia, and Lamberto Ballan. Social and scene-aware trajectory...


