arXiv:2302.07583v1 [cs.CV] 15 Feb 2023

# ForceFormer：探索社会力与 Transformer 在行人轨迹预测中的应用

张伟成[1], 程昊[2],Fatema T. Johora[3] 和 Monika Sester[1]

摘要——**在交互频繁的场景中，基于目标信息预测行人轨迹，是迈向智能交通系统和自动驾驶的关键一步。这项任务的挑战主要来自两个方面：（1）行人密度较高时复杂的社会交互；（2）对目标信息的利用有限，难以有效地将其与过往运动信息关联起来。为应对这些难题，我们将社会力融入 Transformer 随机生成模型骨干网络，并提出一种新型基于目标的轨迹预测器，称为 *ForceFormer*。不同于大多数仅将目的地位置用作输入特征的先前研究，我们利用目的地产生的驱动力，高效模拟目标对行人的引导作用。此外，我们将排斥力作为另一种输入特征，用于描述相邻行人之间的避让行为。大量实验表明，在常用行人数据集上，我们提出的方法在距离误差方面达到与最先进模型相当的性能，同时显著减少了碰撞，尤其是在行人密集场景中。**

## I. 引言

在人群场景中准确且合理地预测行人轨迹，对多种应用都至关重要，例如移动机器人导航 [1], 智能交通系统和智能车辆 [2]，以及共享空间安全 [3]。与受交通规则（如车道几何、交通信号灯和行进方向）约束的车辆运动不同，行人可能随时停下或转向，并且与邻居的互动更多，因此其行为具有高度随机性。

为了对行人行为进行建模，人们采用了多种不同的方法。在基于规则的模型中，行人之间的交互，即社会交互，被描述为力 [4]。在数据驱动模型中，注意力机制 [5] 和图卷积网络 [6] 被广泛用于提取社会交互，并通过监督学习取得出色的结果 [7], [8], [9]。目标信息也可以减少对行人行为的不确定性。然而，这些方法各有缺陷。尽管基于规则的模型能够模拟一定程度的行人行为，但它们的稳健性较差，在复杂场景中缺乏韧性。数据驱动模型通常能取得更好的性能，但依赖数据且可解释性较差 [10]。在基于目标的模型中，目标信息通常直接作为输入，或作为从当前位置到目标的偏移量 [11], [12], [13]。这类模型并非最优，因为目标与当前位置之间尚未建立关联。

为此，我们提出了一种新颖的基于目标的行人轨迹预测框架，称为 *ForceFormer*。它不仅以序列运动信息作为输入，还利用力来训练基于Transformer的主干网络。与以往模型直接将最后位置的信息作为输入特征、并与描述运动动力学的其他特征并列使用不同，我们利用目标信息推导社会力，从而使速度、位置和方向的变化与目标信息更紧密地关联起来。

此外，我们采用生成模型 AgentFormer [9]作为轨迹预测骨干网络，它利用 Transformer [5] 网络学习时间维度上的社会交互。同时，为了在推理时估计用于计算力的临时目标位置，我们采用了目标估计模块 [14], [13] 。具体而言，我们将历史轨迹与语义场景信息拼接起来，并将其输入 U-Net [15]结构以预测潜在目标。借助这一目标估计模块，我们能够获得可靠的目标信息，同时自然地考虑环境因素的约束。

总而言之，我们的主要贡献如下：

- 我们提出了一种基于目标的轨迹预测框架 **ForceFormer**。该框架将更具可解释性的特征，即社会力，引入数据驱动模型，以学习随机行人行为。
- 我们研究了多种利用目标信息的 ForceFormer 变体，并发现其中两种效果显著。具体而言，a)**ForceFormer-Re** 利用目标位置推导排斥力，强化自我行人与邻居之间的交互信息，并降低发生碰撞的可能性；b)**ForceFormer-Dr** 利用目标推导驱动力，增强目的地引导，以预测自我行人的未来轨迹。
- 我们在广泛使用的 ETH/UCY [16], [17] 行人数据集上进行了大量实证研究。实验结果表明，ForceFormer 在标准距离误差指标上与最先进模型表现相当，但能显著减少碰撞，尤其是在行人密集场景中。

[1]张伟成和Monika Sester任职于德国汉诺威莱布尼茨大学制图与地理信息学研究所，地址为Appelstr. 9, 30167 Hannover, Germanyweicheng.zhang@stud.uni-hannover.de, Monika.Sester@ikg.uni-hannover.de[2]程昊任职于荷兰特温特大学地理信息科学与地球观测学院，地址为7500 AE Enschede, The Netherlandsh.cheng-2@utwente.nl[2]Fatema T. Johora任职于德国克劳斯塔尔工业大学信息学系，地址为Julius-Albert-Str. 4, 38678 Clausthal-Zellerfeld, Germany fatema.tuj.johora@tu-clausthal.de

## II. 相关工作

本节简要回顾序列建模、社会交互建模和基于目标的模型相关研究。

**序列建模**。本质上，运动轨迹由时间序列中的位置信息构成。因此，将轨迹预测转化为序列到序列建模是最常用的方法之一。在以往的研究中，得益于强大的门控函数，长短期记忆网络（LSTM） [18] 已广泛应用于许多行人轨迹预测任务，并取得了出色的结果，尤其是在时间维度上 [7], [6],[19], [20], [8]。近年来，Transformer [5] 网络在自然语言处理（NLP）领域取得了巨大成就 [21], [22], ，基于Transformer的模型也被应用于轨迹预测。与LSTM相比，Transformer网络基于自注意力机制，在对长序列中的时间依赖关系进行建模方面能力更强 [23], [24], [9]。除了LSTM等传统的确定性方法外，越来越多的深度生成模型，如条件变分自编码器（CVAEs） [25],[26] 和生成对抗网络（GANs） [27] ，也被用于轨迹预测。生成模型并非只生成单一预测，而是将潜在的未来轨迹学习为一个分布，并从潜在空间生成多个可能的预测。例如，Social GAN和Sophie [28], [29] 通过联合训练生成器和判别器，被提出用于行人的多路径轨迹预测。与GAN相比，条件变分自编码器模型以过去轨迹为条件预测多条合理轨迹，并在近期研究中取得了更好的性能 [30], [31], [32], [8], [33]。

**社会交互建模**。除了对个体轨迹序列进行建模，刻画行人彼此之间以及行人与环境之间的影响，一直是行人轨迹预测中的关键问题。作为开创性工作，Helbing等人 [4] 利用动态社会力来模拟周围环境对行人的影响，例如用于碰撞规避的排斥力和用于建立社会联系的引力。社会力模型已有效应用于机器人学 [34] 和人群分析 [35], [36]等多个领域。另一项开创性的数据驱动工作是Social LSTM [7]。该方法提出了一种称为社会池化层的新结构，用于聚合来自邻居的交互信息。随着图神经网络（GNN） [37], 的发展，近期更多确定性模型的相关工作，如 [24], [38] ，将人群建模为图，并结合 GNN 与注意力机制来学习空间交互。其他方法，如 [32], [6] ，则首先在每个独立时间步对社会维度上的特征进行编码，然后将这些社会特征输入另一个时间序列模型，以概括随时间变化的社会关系。与上述方法不同，我们在每个时间步将社会力引入基于Transformer的主干网络，从而促进对行人之间社会交互的学习。

**基于目标的模型**。近年来，基于目标的模型已成为提升预测性能的有效方法[11], [39]。多样化的目标信息能够为确定性模型提供更多预测可能性 [40], [13]。此外，行人的行动受其目的地驱动。因此，目标信息 [14]可以减少行为中的高度不确定性。不同于直接将目标信息用作输入特征的模型，我们利用目标信息计算每位行人的社会力 [4], ，并将所得的力作为预测模块的输入。然而，在推理时无法获取目标信息。为解决这一问题，我们使用目标估计 [14] 模块在推理时估计目标。

## III. 方法

### A. 问题表述

在行人轨迹预测问题中，行人的完整轨迹可分为两部分：观测轨迹和未来轨迹。时间步 $t \leq 0$ 处的观测轨迹记为 $X = (X^{-H}, X^{-H+1}, ...X^{0})$，共包含 $H+1$ 个观测时间步；而未来 T 个时间步内时间步 $t > 0$处的未来轨迹记为 $Y = (Y^{1}, Y^{2}, ..., Y^{T})$。与 [9], 类似，我们使用二维坐标系中的 x、y 坐标和速度序列来参数化轨迹。此外，同一场景中同一时间步 t 的所有 N 个行人的联合社会序列，观测部分记为 $X^{t} = (x_1^t, x_2^t, ..., x_N^t)$，未来轨迹部分记为 $Y^{t} = (y_1^t, y_2^t, ..., y_N^t)$。在我们提出的生成模型 $p_\theta(Y|X, G, F)$ 中，其中 $\theta$为模型参数，任务是根据未来轨迹 Y 进行预测，预测依据不仅包括观测轨迹 X，还包括目标信息 G 和社会力 F。

接下来 [13],我们使用最后一个时间步的位置以及每个位置与目标位置之间的差值来参数化目标信息。需要注意的是，在训练阶段，我们使用最后一个位置的真实值$Y^T$ 来推导目标表示；而在测试阶段，我们使用第 III-B 节目标估计模块的估计结果来推导目标表示。此外，每个智能体在每个时间步都会计算两种力，即驱动力$F_{\mathrm{Dr}}$ 和排斥力$F_{\mathrm{Re}}$。它们也表示为序列。

$$F = \begin{cases} F_{\mathrm{Dr}} = (f_{\mathrm{Dr}_1}^{-H}, ..., f_{\mathrm{Dr}_N}^{-H}, ..., f_{\mathrm{Dr}_1}^{T}, ..., f_{\mathrm{Dr}_N}^{T}), \\ F_{\mathrm{Re}} = (f_{\mathrm{Re}_1}^{-H}, ..., f_{\mathrm{Re}_N}^{-H}, ..., f_{\mathrm{Re}_1}^{T}, ..., f_{\mathrm{Re}_N}^{T}). \end{cases} \quad (1)$$

为了探索融入目标信息的不同方式，我们以基线模型 Agent- Former [9] 为基础，该模型将速度和位置序列作为输入，并提出了三种加入额外目标信息的变体。如图 1 所示，**ForceFormer-Goal** 直接将额外的目标序列添加到输入中。另一种方式是，**ForceFormer-Dr**使用$F_{\mathrm{Dr}}$ 作为额外的条件信息。**ForceFormer-Re** 则将目标序列和排斥力$F_{\mathrm{Re}}$ 序列同时作为额外输入。

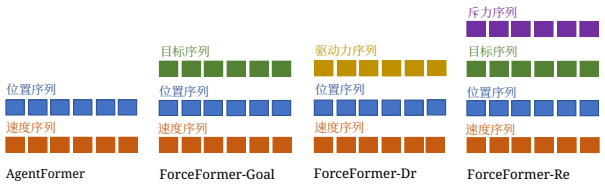


图 1. AgentFormer [9] 和所提出的 ForceFormer-Goal、ForceFormer-Dr 和 ForceFormer-Re 的输入。

### B. 所提出的框架

图 2 展示了我们提出的框架 ForceFormer 的整体结构。它主要由三个模块组成：作为骨干网络的 AgentFormer（X-编码器、Y-编码器和解码器）、目标估计模块以及社会力模型。

在训练过程中，目标估计模块和 AgentFormer 分别进行训练。目标信息由真实值 $Y^T$ 提供，用于训练目标估计模块并计算社会力。排斥力和驱动力根据位置、速度和目标信息计算。然而，在测试阶段无法获取真实值 $Y^T$ 。因此，在推理过程中，我们从目标估计模块为每条轨迹采样 $K$ 个目标候选。按照以往基于目标条件的轨迹预测模型 [39], [11], [14],，我们将所有可能的$K$ 个目标与真实值进行比较，并选择$L2$误差最小的目标，作为测试阶段估计的目标位置。

*a) AgentFormer*： 骨干预测模型采用基于CAVE的模型，并通过注意力机制建立空间和时间关系。基于观测轨迹 $X$、目标信息 $G$和社会力 $F$这些条件，对未来轨迹分布建模为 $p_\theta(Y|X,\ G,\ F)$。未来轨迹分布可改写为

$$p_\theta(Y|X,G,F)=\int p_\theta(Y|Z,X,G,F)p_\theta(Z|X,G,F)dZ, \quad (2)$$

其中，$p_\theta(Z|X,G,F)$ 是条件高斯先验，由 X-编码器学习得到。$p_\theta(Y|Z,X,G,F)$ 是条件似然。公式（2）提出了一组潜变量$Z=(z^{(1)},...,z^{(K)})$，用于反映行人$n$的潜在意图，以刻画行人未来行为的随机性和多模态性。

负证据下界 $\mathscr{L}_{elbo}$ 用于处理不可处理的后验分布$p_\theta(Z|Y,X,G,F)$。具体而言，基于CVAE的模型通过以下损失函数进行优化：

$$\begin{aligned}\mathscr{L}_{elbo}=&-\mathbb{E}_{q_\phi(Z|Y,X,G,F)}[\log p_\theta(Y|Z,X,G,F)]\\&+KL(q_\phi(Z|Y,X,G,F)||p_\theta(Z|X,G,F)),\end{aligned} \quad (3)$$

其中，$q_\phi(Z|Y, X, G,F)$ 是由 $\phi$ 参数化的近似后验分布，该参数通过 Y-Encoder 学习得到。上述公式中的第一项可视为未来轨迹 $p_\theta(Y|Z,X,G, F)$ 的期望预测概率。第二项 $KL(q_\phi(Z|Y, X, G,F)||p_\theta(Z|X,G, F))$ 表示该分布先验分布与近似后验分布之间的差异，而这两个分布都趋近于标准正态分布。

*b)* 社会力：本文采用了一种改进的社会力模型 [4]，其中包含驱动力和排斥力。式（4）所示的**驱动力**描述了与目的地（目标位置）相关的吸引作用。

$$\vec{F}_\alpha^0=\frac{1}{\tau_\alpha}(v_\alpha^0\vec{e}_\alpha-\vec{v}_\alpha). \quad (4)$$

驱动力的大小取决于当前速度 $\vec{}v\alpha(t)$ 与期望速度$v_\alpha^0(t)=v_\alpha^0\vec{}e\alpha$之间的偏差。速度$v_\alpha^0(t)=v_\alpha^0\vec{}e\alpha$ 表示，如果行人未受到干扰，她将以期望速度$v_\alpha^0$ 沿着指向目的地的期望方向 $\vec{}e\alpha$ 行走。松弛时间 $\tau_\alpha$ 是一个参数，表示消除该偏差所需的预期时间。

为避免碰撞，行人会与其他陌生人保持适当距离。式 5 所表示的**排斥力**描述了自我行人 $\alpha$ 与其他行人 $\beta$ 之间的避让现象，

$$\vec{f}_{\alpha\beta}(\vec{r}_{\alpha\beta})=-\nabla_{\vec{r}_{\alpha\beta}}V_{\alpha\beta}[b(\vec{r}_{\alpha\beta})]. \quad (5)$$

排斥势能 $V\alpha\beta(b)$ 是关于$b$的单调递减函数，其中 $b$ 表示椭圆的半短轴。通过$b$，等势线保持椭圆形状，并指向运动方向。

$$2b=\sqrt{(\|\vec{r}_{\alpha\beta}\|+\|\vec{r}_{\alpha\beta}-v_\beta\Delta t\vec{e}_\beta\|)^2-(v_\beta\Delta t)^2}. \quad (6)$$

计算排斥力时，除了距离之外，还需要考虑视角的影响。因此，引入参数$w$ 。

$$w(\vec{e},\vec{f})=\begin{cases}1 & \text{if } \vec{e}\cdot\vec{f}\geq\left\|\vec{f}\right\|cos\varepsilon,\\ c & \text{otherwise,}\end{cases} \quad (7)$$

其中，有效视角为 $2\varepsilon$，$c$ 为常数因子。因此，考虑视角因子后的排斥力受到如下约束： $\vec{}F\alpha\beta=w\vec{}f\alpha\beta$。

由于排斥力基于这样一个前提：两名行人互不相识，因此他们希望彼此保持距离并避免碰撞。然而，实际上，许多行人会结伴出行，例如同学、亲属和朋友，他们有共同目的地。因此，将群组内的排斥力纳入考虑并不合理，这可能会导致实验结果出现显著误差，尤其是在行人密度较高的场景中。于是，我们在每个时间步采用 DBSCAN 方法，即具有噪声的基于密度的空间聚类 [41], [42] 。通过聚类检测潜在的群组成员 [43]。也就是说，如果两名行人在观测帧中有超过 $\sigma$ 个时间步处于同一簇中，就会将他们判定为同一群组。随后，消除群组内排斥力。

*c)* 目标估计模块：目标估计模块用于为 AgentFormer 预测模块提供目标信息，并在推理阶段计算社会力。为此，我们采用目标模块

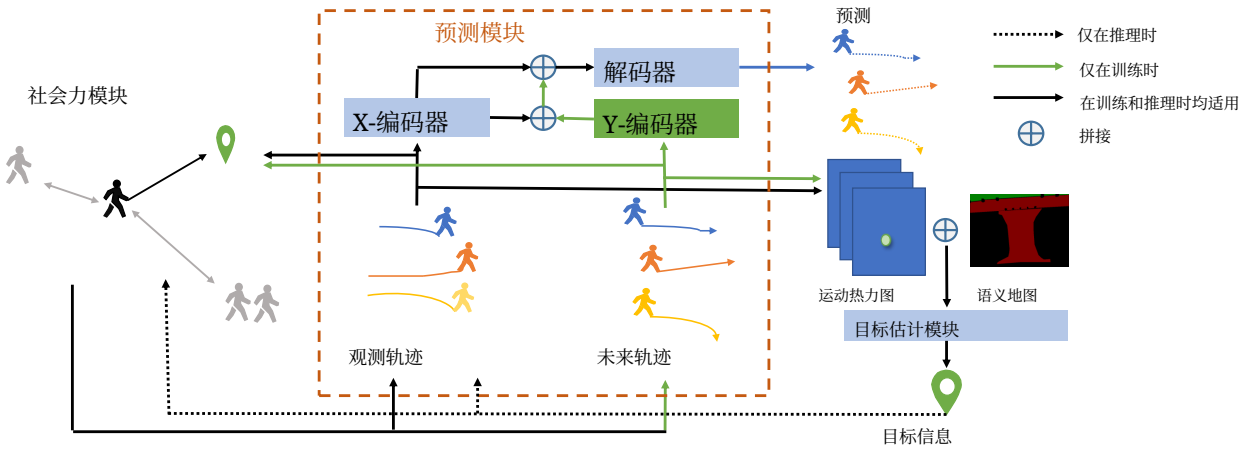


图 2. 所提出的 ForceFormer 框架概览。

在 [14], [13] 中提出的方法。首先，将热力图形式的过去轨迹与语义地图信息进行拼接。语义地图采用 Chiara 等人 [13],的方法，通过预训练分割网络 [14]从 ETH/UCY 数据集中的 RGB场景图像鸟瞰视图中提取。具体而言，借助语义地图，可以自然地考虑环境约束，例如路面、地形和建筑物。分割结果以张量形式表示为 $S \in \mathbb{R}^{W\times L\times C}$，其中包含$C$ 个类别。$H$ 和 $L$ 分别表示输入图像的高度和宽度。智能体$n$的过去轨迹$\{x_n^{-H}, x_n^{-H+1}, \ldots, x_n^{0}\}$ 被映射为热力图 $M \in \mathbb{R}^{W\times L\times (H+1)}$。随后，沿通道维度将过去轨迹的热力图张量与语义地图 $S$ 拼接，生成张量 $M_s \in \mathbb{R}^{W\times L\times (C+H+1)}$ ，作为目标估计的输入张量。最后，将拼接后的信息输入 U-Net [15] 模型，以生成未来位置的概率图。

## IV. 实验

### A. 数据集

所提出的框架在 ETH [16]和 UCY [17], 数据集上进行了评估，这些数据集已被广泛用作行人轨迹预测的基准。如表 I 所示，这些数据集包含五个不同的子集。有效轨迹指在以 2.5 Hz 采集的连续 20 帧中，单个行人的轨迹信息。这 20 个时间步分为两部分：前 8 个时间步（3.2 秒）为观测轨迹 $X$，基于这些观测轨迹，预测后续 12 个时间步（ 4.8 秒）的未来轨迹 $Y$ 。目标的位置位于第 20 个时间步。可以看出，不同子集中的行人密度有所不同。场景中的密度会显著影响预测结果，尤其是在本研究中，因为社会力的计算与人群密度密切相关。

### B. 最先进模型和基线

我们将提出的方法 ForceFormer 与以下模型进行比较。AgentFormer [9]是不使用任何目标信息的基线模型

表 I ETH/UCY[16], [17]数据集中的帧数和有效轨迹数。

| | ETH | Hotel | Univ | zara01 | zara02 |
|---|---|---|---|---|---|
| 帧 | 1142 | 1788 | 947 | 883 | 1033 |
| 有效轨迹 | 364 | 1197 | 24334 | 2356 | 5910 |

。Sophoie [29] 提出了一种基于生成对抗网络的模型，将轨迹信息与上下文信息相结合。Trajectron++ [32] 是一种基于条件变分自编码器的模型，在 ETH/UCY 基准上始终保持出色表现。STAR [24]提出了时间 Transformer 和空间 Transformer，用于对行人轨迹预测中的时空信息进行建模。此外，ForceFormer 还与一系列基于目标的模型进行了比较。具体而言，PECNet [11] 是一种用于短期轨迹预测的目标条件模型。Goal-GAN [12] 将目标信息融入基于生成对抗网络的轨迹预测模型中。Heading [39] 提出了一个目标检索模块，为轨迹预测提供目标信息。Y-net [14] 将场景信息与目标和航点相结合，用于轨迹预测。Goal-SAR [13] 提出了一种基于注意力的循环网络，并结合了与 ForceFormer 相同的目标估计模块。

### C. 评估指标与协议

本文使用三个指标评估所提出的模型。首先，采用两个标准误差指标来衡量 ForceFormer 的轨迹预测性能，并与以往模型进行公平比较。这两个距离误差分别是平均位移误差$ADE_K$ 和最终位移误差$FDE_K$ ，用于衡量每个智能体的$K$ 个轨迹样本与相应真实值之间的差异。

$$ADE_K = \frac{1}{T}\min_{k=1}^{K}\sum_{t=1}^{T}\left\|\hat{y}_n^{t,(k)} - y_n^t\right\|^2, \tag{8}$$

$$FDE_K = \min_{k=1}^{K}\left\|\hat{y}_n^{T,(k)} - y_n^T\right\|^2. \tag{9}$$

此外，还利用碰撞次数这一指标来验证应用于 ForceFormer 的社会力。

$$NC = \sum_{m,n=1}^{N} \sum_{t=1}^{T} (\left\| \hat{y}_n^t - \hat{y}_m^t \right\|) < \gamma, m \neq n, \tag{10}$$

其中，$m$ 和$n$ 是时间步$t$同一场景中的不同行人。设置阈值 $\gamma$以判断是否发生碰撞。在本文中， $\gamma = 0.1m$。所有指标均使用$K = 20$ 个样本计算。$NC$ 的计算基于平均位移误差最小的轨迹。遵循以往研究 [44], [9], [32], [11], , 我们采用留一法进行评估。

### *D.* 实现细节

在计算社会力时，我们采用 200° 作为有效角度 $2\varepsilon$。视野因子$c$ 为 0.5。分组 $\sigma$ 时，同一簇中的最小帧数阈值设为 4。此外，我们认为当行人位置与目标位置重叠时，无法计算期望方向，因此将这些位置上的社会力设为零。对于 AgentFormer 主干网络，我们采用与 Yuan 等人 [9] 相同的所有设置。但我们仅使用 Adam 优化器 [45]训练条件变分自编码器模型 50 个训练轮次，少于原论文中的轮次。对于目标估计模块，除采用与 Chiara 等人相同的设置外，[13], 我们还加入了特定于目标的均方误差损失函数 $\mathscr{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} \left\| y_i^T - \hat{y}_i^T \right\|^2$ 以及超参数 $\lambda = 1e^6$，以平衡原有的二元交叉熵损失函数。我们所有的模型均在 Google Colab 上使用单块 Tesla P100 GPU 训练。

### *E.* 结果

在表 II 中，我们将提出的方法与当前最先进的方法进行了比较。首先，与基线模型 AgentFormer 相比，我们提出的方法 ForceFormer-Dr 和 ForceFormer-Re 在所有子集上都取得了更好的性能。例如，平均而言，ForceFormer-Dr 将最终位移误差降低了 26%，ForceFormer-Re 将平均位移误差降低了 17%。此外，与同样使用目标信息的模型相比，我们的方法与此前表现最佳的方法 Y-net 不相上下。具体而言，比较各子集上的结果可以发现，除 ETH 外，我们的模型在其余四个子集上的表现都优于 Y-net。此外，在高密度场景中，即 Univ 和 Zara2 上，与 Y-net 相比，ForceFormer-Dr 将最终位移误差分别降低了 22% 和 23%。这些改进表明，我们的方法在高密度场景中的最终位置预测方面表现更佳。

### *F.* 消融研究

表 III 对比了我们提出的利用目标信息的模型变体的 FDE 值，表 IV 则对比了碰撞次数。首先可以清楚地看到，与基线模型 AgentFormer 相比，所有利用额外目标信息的变体都取得了更低的平均 FDE。除 ForceFormer-Goal 外，ForceFormer-Dr 和 ForceFormer-Re 的碰撞次数也明显更少。在 ForceFormer 的三个变体中，

表 II 最先进模型与我们提出的模型在 ETH/UCY 数据集上的定量表现。

| 方法 | 平均位移误差$_k$/最终位移误差$_k$（米）$K = 20$ 样本 | | | | | |
|---|---|---|---|---|---|---|
| 数据集 | ETH | Hotel | Univ | Zara1 | Zara2 | 平均值 |
| Sophie [29] | 0.70/1.43 | 0.76/1.67 | 0.54/1.24 | 0.30/0.63 | 0.38/0.78 | 0.54/1.15 |
| STAR [24] | 0.36/0.65 | 0.17/0.36 | 0.31/0.62 | 0.26/0.55 | 0.22/0.46 | 0.26/0.53 |
| Trajectron++ [32] | 0.67/1.18 | 0.18/0.28 | 0.30/0.54 | 0.25/0.41 | 0.18/0.32 | 0.32/0.55 |
| AgentFormer [9] | 0.45/0.75 | 0.14/0.22 | 0.25/0.45 | 0.18/0.30 | 0.14/0.24 | 0.23/0.39 |
| PECNet [11] | 0.54/0.87 | 0.18/0.24 | 0.35/0.60 | 0.22/0.39 | 0.17/0.30 | 0.29/0.48 |
| Goal-GAN [12] | 0.59/1.18 | 0.19/0.35 | 0.60/1.19 | 0.43/0.87 | 0.32/0.65 | 0.43/0.85 |
| Heading [39] | 0.37/0.65 | 0.11/0.15 | **0.20**/0.44 | 0.15/0.31 | **0.12**/0.26 | 0.19/0.36 |
| Y-net [14] | **0.28**/**0.33**0.10/**0.14**0.24/0.41 | | 0.17/0.27 | 0.13/0.22 | **0.18**/**0.27** | |
| Goal-SAR [13] | **0.28**/0.39 | 0.12/0.17 | 0.25/0.43 | 0.17/0.26 | 0.15/0.22 | 0.19/0.29 |
| ForceFormer-Dr | 0.43/0.58 | 0.12/0.16 | 0.21/**0.32** | **0.14**/**0.20** | **0.12**/**0.17**0.20/0.29 | |
| ForceFormer-Re | 0.36/0.52 | **0.09**/**0.14**0.21/0.42 | | 0.15/0.22 | **0.12**/0.20 | 0.19/0.30 |

*根据实现问题 53 [46] 和采样技巧 [47]，更新了 Trajectron++ 和 Heading 的结果。带下划线的方法使用了目标信息。

表 III 使用目标信息的方法在最终位移误差方面的性能差异。

| | 最终位移误差$_k$(米)$K = 20$ 样本 | | | | | | |
|---|---|---|---|---|---|---|---|
| 方法 | 目标 | 最$_{Dr}$ | 终$_{Re}$ | ETH | Hotel | Univ | Zara1 | Zara2 | 平均 |
| AgentFormer [9] | | | | 0.75 | 0.22 | 0.45 | 0.30 | 0.24 | 0.39 |
| ForceFormer-Goal | ✓ | | | 0.55 | 0.17 | 0.49 | 0.30 | 0.28 | 0.36 |
| ForceFormer-Dr | | ✓ | | 0.580.16 | | **0.32** | **0.20** | **0.17** | **0.29** |
| ForceFormer-Re | ✓ | | ✓ | 0.52**0.14** | | 0.42 | 0.22 | 0.20 | 0.30 |

总体而言，ForceFormer-Goal 在各个子集上的 FDE 和碰撞次数均逊于其他模型。这表明，直接利用目标序列可能不如利用社会力有效。仔细观察可发现，在 Univ 和 Zara02 等高密度行人场景中，ForceFormer-Dr 的 FDE 最低。相比之下，ForceFormer-Re 的碰撞总数最少，与基线模型 AgentFormer 相比减少了 19.8%。

表 IV 使用碰撞次数衡量采用目标信息的方法之间的性能差异。

| | **碰撞次数** $CN_K, K = 20$ **样本** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **方法** | 目标 $F_{Dr}$ | $F_{Re}$ | ETH | Hotel | Univ | Zara1 | Zara2 | 总计 |
| AgentFormer [9] | | | 0 | 2 | 655 | 4 | 22 | 683 |
| ForceFormer-Goal | ✓ | | **0** | **0** | 672 | **3** | 22 | 697 |
| ForceFormer-Dr | | ✓ | **0** | 1 | 556 | **3** | 28 | 588 |
| ForceFormer-Re | ✓ | ✓ | **0** | 1 | **529** | 5 | **13** | **548** |

### *G.* 定性结果

图3分别展示了 AgentFormer（左列）、ForceFormer-Dr（中列）和 ForceFormer-Re（右列）预测的定性结果。从上排可以看出，与 AgentFormer 相比，ForceFormer-Dr 利用目标信息和驱动力，能够预测行人在拐角或转弯处的轨迹。尽管 ForceFormer-Re 预测的弯曲轨迹不够准确，但其对其他轨迹的预测更接近

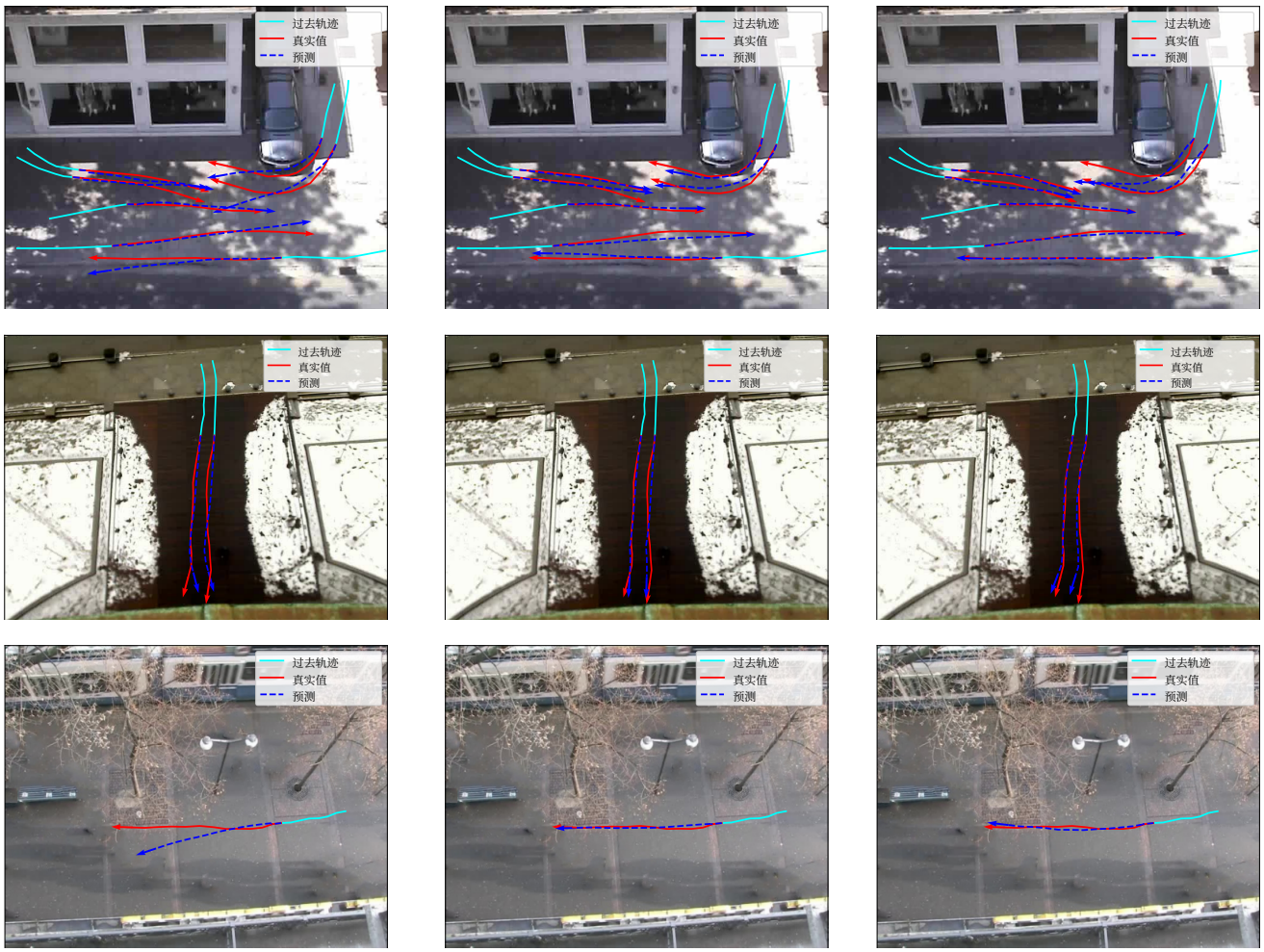


图 3. AgentFormer（左列）、ForceFormer-Dr（中列）和 ForceFormer-Re（右列）生成的预测结果。每一行代表一个不同的场景。

相应的真实值。对于中排中两名行人并行行走的场景，当行人左转时，ForceFormer-Dr 和 ForceFormer-Re 都能更准确地预测最终位置。相比之下，AgentFormer 没有利用目标估计模块中的目标信息，而是预测行人在道路中间行走。在底排中可以看到一个预测最终位置的更直观场景。AgentFormer 的预测与真实值轨迹存在较大偏差，而 ForceFormer-Dr 和 ForceFormer-Re 的预测则与真实值轨迹高度吻合。

**局限性**。尽管社会力和目标估计模块带来了性能提升，但仍需注意所提出模型存在若干局限。碰撞有所减少，但 ForceFormer 的预测并未完全避免碰撞。一个可能的原因是，在社会力模块中，我们没有考虑群体内部的交互和力，而这些因素同样可能导致碰撞。在未来的工作中，我们将构建一个更全面的社会力模块，并将其应用于更好地模拟群组成员之间的交互。此外，ForceFormer 的整体性能，尤其是社会力的计算，依赖于目标估计模块的可靠性。该模块的次优性能可能会导致最终预测中的误差累积。另一方面，如果我们能够获取真实值目标信息，就可以迅速将我们的模型转变为运动规划模型。

## V. 结论

本文提出了一种名为 ForceFormer 的新型基于目标的轨迹预测器，其将社会力融入基于 Transformer 的生成模型骨干网络中。采用基于 U-Net 的目标估计模块来预测行人轨迹的目标。除位置和速度信息外，我们还根据估计的目标推导出驱动力，从而高效模拟目标对行人的引导作用。此外，模型利用排斥力帮助学习相邻行人之间的碰撞规避。ForceFormer 的性能与当前最先进的模型相当，并且在广泛使用的行人数据集上的高密度场景中表现更佳。

## 参考文献


[1] Y. Luo、P. Cai、A. Bera、D. Hsu、W. S. Lee 和 D. Manocha，“Porca：众多行人环境下的自动驾驶建模与规划”，IEEE机器人与自动化快报，第3卷，第4期，第3418–3425页，2018年。[2] L. Ballan、F. Castaldo、A. Alahi、F. Palmieri 和 S. Savarese，“面向特定场景运动预测的知识迁移”，载于《欧洲计算机视觉会议》论文集。施普林格，2016年，第697–713页。[3] Y. Li、H. Cheng、Z. Zeng、H. Liu 和 M. Sester，“自动驾驶车辆驶入共享空间：聚焦弱势道路使用者的外部人机界面设计概念”，载于《2021年IEEE国际智能交通系统会议（ITSC）》。IEEE，2021年，第1729–1736页。[4] D. Helbing 和 P. Molnar，“行人动力学的社会力模型”，《物理评论E》，第51卷，第5期，第4282页，1995年。[5] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、Ł. Kaiser 和 I. Polosukhin，“注意力机制即一切”，《神经信息处理系统进展》，第30卷，2017年。[6] Y. Huang、H. Bi、Z. Li、T. Mao 和 Z. Wang，“STGAT：用于人体轨迹预测的时空交互建模”，载于《IEEE/CVF 计算机视觉与模式识别会议》论文集，2019年，第6272–6281页。[7] A. Alahi、K. Goel、V. Ramanathan、A. Robicquet、L. Fei-Fei 和 S. Savarese，“Social LSTM：拥挤空间中的人体轨迹预测”，载于《IEEE/CVF 计算机视觉与模式识别会议》论文集，2016年，第961–971页。[8] H. Cheng、W. Liao、M. Y. Yang、B. Rosenhahn 和 M. Sester，“AMENet：用于轨迹预测的注意力地图编码器网络”，《国际摄影测量与遥感杂志》，第172卷，第253–266页，2021年。[9] Y. Yuan、X. Weng、Y. Ou 和 K. M. Kitani，“AgentFormer：面向社会时空多智能体预测的智能体感知 Transformer”，载于《IEEE/CVF 计算机视觉与模式识别会议》论文集，2021年，第9813–9823页。[10] H. Cheng、F. T. Johora、M. Sester 和 J. P. M¨uller，“共享空间中的轨迹建模：基于专家的方法与深度学习方法之比较？”，载于《多智能体系统与基于智能体仿真国际研讨会》。施普林格，2021年，第13–27页。[11] K. Mangalam、H. Girase、S. Agarwal、K.-H. Lee、E. Adeli、J. Malik 和 A. Gaidon，“重要的不是旅程，而是目的地：端点条件轨迹预测”，载于《欧洲计算机视觉会议》论文集。施普林格，2020年，第759–776页。[12] P. Dendorfer、A. Osep 和 L. Leal-Taix´e，“Goal-GAN：基于目标位置估计的多模态轨迹预测”，载于《亚洲计算机视觉会议》论文集，2020年。[13] L. F. Chiara、P. Coscia、S. Das、S. Calderara、R. Cucchiara 和 L. Ballan，“用于轨迹预测的目标驱动自注意力循环网络”，载于《IEEE/CVF 计算机视觉与模式识别会议》论文集，2022年，第2518–2527页。[14] K. Mangalam、Y. An、H. Girase 和 J. Malik，“从目标、航点和路径到长期人体轨迹预测”，载于《IEEE/CVF 计算机视觉与模式识别会议》论文集，2021年，第15 233–15 242页。[15] O. Ronneberger、P. Fischer 和 T. Brox，“U-Net：用于生物医学图像分割的卷积网络”，载于《医学图像计算与计算机辅助干预国际会议》。施普林格，2015年，第234–241页。[16] S. Pellegrini、A. Ess、K. Schindler 和 L. Van Gool，“你永远不会独行：面向多目标跟踪的社会行为建模”，载于《IEEE/CVF 计算机视觉与模式识别会议》论文集。IEEE，2009年，第261–268页。[17] A. Lerner、Y. Chrysanthou 和 D. Lischinski，“ 以实例为范例的人群 ”，载于《计算机图形论坛》，第 26 卷，第3期。Wiley在线图书馆，2007年，第655–664页。[18] S. Hochreiter 和 J. Schmidhuber，“长短期记忆”，《神经计算》，第9卷，第8期，第1735–1780页，1997年。[19] P. Zhang、W. Ouyang、P. Zhang、J. Xue 和 N. Zheng，“SR-LSTM：面向行人轨迹预测的 LSTM 状态细化”，载于《IEEE/CVF 计算机视觉与模式识别会议》论文集，2019年，第12 085–12 094页。[20] Y. Xu、Z. Piao 和 S. Gao，“利用深度神经网络编码人群交互以进行行人轨迹预测”，载于《IEEE/CVF 计算机视觉与模式识别会议》论文集，2018年，第5275–5284页。[21] J. Devlin、M.-W. Chang、K. Lee 和 K. Toutanova，“BERT：用于语言理解的深度双向Transformer预训练”，arXiv预印本 arXiv:1810.04805，2018年。[22] Z. Lan、M. Chen、S. Goodman、K. Gimpel、P. Sharma 和 R. Soricut，“ALBERT：用于语言表示自监督学习的轻量级 BERT”，arXiv预印本 arXiv:1909.11942，2019年。[23] F. Giuliari、I. Hasan、M. Cristani 和 F. Galasso，“用于轨迹预测的 Transformer网络”，载于《2020年第25届国际模式识别会议（ICPR）》。IEEE，2021年，第10 335–10 342页。

[24] C. Yu、X. Ma、J. Ren、H. Zhao 和 S. Yi，“用于行人轨迹预测的时空图 Transformer 网络”，载于《欧洲计算机视觉会议论文集》。施普林格，2020年，第 507–523 页。[25] D. P. Kingma 和 M. Welling，“自编码变分贝叶斯”，arXiv 预印本 arXiv:1312.6114，2013年。[26] K. Sohn、H. Lee 和 X. Yan，“使用深度条件生成模型学习结构化输出表示”，《神经信息处理系统进展》，第 28 卷，2015年。[27] I. Goodfellow、J. Pouget-Abadie、M. Mirza、B. Xu、D. Warde-Farley、S. Ozair、A. Courville 和 Y. Bengio，“生成对抗网络”，《ACM 通讯》，第 63 卷，第 144 页，2020年11月，第 139 页。[28] A. Gupta、J. Johnson、L. Fei-Fei、S. Savarese 和 A. Alahi，“Social GAN：利用生成对抗网络生成社会可接受轨迹”，载于 IEEE/CVF 计算机视觉与模式识别会议论文集，2018年，第 2255–2264 页。[29] A. Sadeghian、V. Kosaraju、A. Sadeghian、N. Hirose、H. Rezatofighi 和 S. Savarese，“Sophie：一种用于预测符合社会与物理约束路径的注意力生成对抗网络”，载于 IEEE/CVF 计算机视觉与模式识别会议论文集，2019年，第 1349–1358 页。[30] B. Ivanovic 和 M. Pavone，“Trajectron：利用动态时空图进行概率多智能体轨迹建模”，载于 IEEE/CVF 计算机视觉与模式识别会议论文集，2019年，第 2375–2384 页。[31] C. Tang 和 R. R. Salakhutdinov，“多未来预测”，《神经信息处理系统进展》，第 32 卷，2019年。[32] T. Salzmann、B. Ivanovic、P. Chakravarty 和 M. Pavone，“Trajectron++：利用异构数据进行动态可行的轨迹预测”，载于《欧洲计算机视觉会议论文集》。施普林格，2020年，第 683–700 页。[33] G. Chen、J. Li、N. Zhou、L. Ren 和 J. Lu，“通过分布判别实现个性化轨迹预测”，载于 ICCV 论文集，2021年10月，第 15 580–15 589 页。[34] G. Ferrer、A. Garrell 和 A. Sanfeliu，“伴侣机器人：一种基于社会力并具有人类感知导航能力、适用于拥挤环境的方法”，载于 2013 年 IEEE/RSJ 国际智能机器人与系统会议。IEEE，2013年，第 1688–1694 页。[35] R. Mehran、A. Oyama 和 M. Shah，“使用社会力模型进行异常人群行为检测”，载于 IEEE/CVF 计算机视觉与模式识别会议论文集。IEEE，2009年，第 935–942 页。[36] F. T. Johora、D. Yang、J. P. M¨uller 和 ¨U. ¨Ozg¨uner，“异构共享交通空间中道路使用者运动模型的泛化能力”，《IEEE 智能交通系统汇刊》，第 23 卷，第 12 期，第 23 084–23 098 页，2022年。[37] T. N. Kipf 和 M. Welling，“使用图卷积网络进行半监督分类”，arXiv 预印本 arXiv:1609.02907，2016年。[38] V. Kosaraju、A. Sadeghian、R. Mart´ín-Mart´ín、I. Reid、H. Rezatofighi 和 S. Savarese，“Social-BiGAT：使用 Bicycle-GAN 和图注意力网络进行多模态轨迹预测”，《神经信息处理系统进展》，第 32 卷，2019年。[39] H. Zhao 和 R. P. Wildes，“你要去哪里？利用专家目标示例进行动态轨迹预测”，载于 ICCV 论文集，2021年10月，第 7629–7638 页。[40] H. Zhao、J. Gao、T. Lan、C. Sun、B. Sapp、B. Varadarajan、Y. Shen、Y. Shen、Y. Chai、C. Schmid 等，“TNT：目标驱动的轨迹预测”，arXiv 预印本 arXiv:2008.08294，2020年。[41] M. Ester、H.-P. Kriegel、J. Sander、X. Xu 等，“一种基于密度的算法：在含噪的大型空间数据库中发现簇”，载于 KDD，第 96 卷，第 34 期，1996年，第 226–231 页。[42] E. Schubert、J. Sander、M. Ester、H. P. Kriegel 和 X. Xu，“重新审视 DBSCAN：为何以及如何（仍然）应使用 DBSCAN”，《ACM 数据库系统汇刊》（TODS），第 42 卷，第 21 页，2017年3月，第 1 页。[43] H. Cheng、Y. Li 和 M. Sester，“共享空间中的行人群体检测”，载于 2019 年 IEEE 智能车辆研讨会（IV）。IEEE，2019 年，第 1707–1714 页。[44] H. Cheng、M. Liu、L. Chen、H. Broszio、M. Sester 和 M. Y. Yang，“GATraj：一种基于图与注意力机制的多智能体轨迹预测模型”，arXiv 预印本 arXiv:2209.07857，2022年。[45] D. P. Kingma 和 J. Ba，“Adam：一种随机优化方法”，arXiv 预印本 arXiv:1412.6980，2014年。[46] T. Salzmann、B. Ivanovic、P. Chakravarty 和 M. Pavone，“Trajectron++：利用异构数据进行动态可行的轨迹预测”，https://github.com/StanfordASL/Trajectron-plus-plus/issues/53，2020年。[47] H. Zhao 和 R. P. Wildes，“你要去哪里？利用专家目标示例进行动态轨迹预测”，https://github.com/JoeHEZHAO/expert traj，2021年。