

# 面向多尺度船舶交互建模的动态时空图 Transformer

杨宗霖

June 21, 2025

## Contents

引言	2
背景与动机	2
问题定义与现有方法的局限性	2
本文提出的解决方案与贡献	3
相关工作	3
基于深度学习的轨迹预测	3
时空图神经网络 (ST-GNNs)	4
智能体交互建模	5
海事碰撞风险评估	5
方法论：动态时空图 Transformer	6
问题形式化与框架概览	7
动态海事安全图构建	7
海事智能体感知图注意力编码器	8
多模态轨迹解码器与训练目标	9
实验设计	10
数据集与预处理	10
评估指标	10
基线模型	11
消融研究	12
预期影响与伦理考量	12
可解释性、因果推断与信任	12

鲁棒性与泛化能力	13
伦理考量	13
结论	14
参考文献	15

## Abstract

随着全球航运密度的持续增长以及海上自主水面船舶 (Maritime Autonomous Surface Ships, MASS) 的加速发展, 对复杂海域中船舶的未来轨迹进行精准、可靠且可解释的预测, 已成为保障航行安全、提升交通效率和实现智能航运的关键技术瓶颈 [8]。现有方法在应对这一挑战时存在明显不足: 传统的海事碰撞风险模型, 如基于最近点距离 (Distance at Closest Point of Approach, DCPA) 和到达最近点时间 (Time to Closest Point of Approach, TCPA) 的模型, 虽然在规则上明确, 但其基于线性运动假设, 难以捕捉船舶在真实交互中的非线性机动行为, 且在密集交通流中易产生误报 [7]。另一方面, 直接从陆基交通 (如自动驾驶车辆或行人) 领域迁移的通用人工智能轨迹预测模型, 虽然具备强大的非线性建模能力, 但它们通常忽略了船舶独特的动力学特性 (如高惯性、慢响应) 以及由《国际海上避碰规则》(COLREGs) 所定义的、基于风险而非距离的隐性交互逻辑, 从而导致领域失配问题 [11]。

为弥合海事领域知识与先进深度学习模型之间的鸿沟, 本研究方案提出一种名为“动态时空图 Transformer (Dynamic Spatio-temporal Graph Transformer, D-STGT)”的新型深度学习框架。D-STGT 的核心创新在于将海事领域的先验安全准则深度嵌入到模型架构的设计中, 实现知识引导的交互建模。我们提出一种由海事安全指标 (TCPA/DCPA) 驱动的动态图构建方法, 该图的拓扑结构及其边权重能够实时、动态地反映船舶间的潜在碰撞风险等级, 而非仅仅依赖于物理上的空间邻近性 [9]。这种设计使得图结构本身成为一种可解释的、符合海事逻辑的风险表征。在此动态风险图的基础上, 我们进一步设计了一种“海事智能体感知图注意力 (Maritime Agent-Aware Graph Attention)”机制。该机制受到 AgentFormer 中代理感知注意力思想的启发 [1], 但创新性地将其从序列模型适配到动态图模型上。它通过并行的注意力路径, 在一个统一的框架内解耦并建模了两种不同尺度的动力学: 单艘船舶自身的时序运动惯性 (Intra-ship Dynamics) 和多艘船舶之间由避碰需求驱动的复杂交互 (Inter-ship Interactions)。

D-STGT 整体采用编码器-解码器架构。海事智能体感知图注意力编码器负责学习动态、多尺度的交互表征, 而解码器则利用条件变分自编码器 (Conditional Variational Autoencoder, CVAE) 的框架生成符合现实物理约束和航行惯例的多模态未来轨迹分布 [5]。我们计划在包括美国国家海洋和大气管理局 (NOAA) 提供的大规模真实船舶自动识别系统 (AIS) 数据集上对模型进行全面验证 [16], 并采用包括联合位移误差 (Joint Displacement Error) 和碰撞率 (Collision Rate) 在内的综合评估指标体系, 以更全面地评估模型的性能和安全性 [10]。预期 D-STGT 将在预测精度、交互真实性和模型可解释性上均超越现有技术, 为开发下一代更安全、更智能的船舶交通管理 (Vessel Traffic Service, VTS) 系统和海上自主导航系统提供关键的技术支撑与理论基础。

## 引言

### 背景与动机

海事运输是全球贸易的生命线，承载了超过 80% 的国际货物贸易量。然而，全球经济一体化进程的不断深化也带来了日益严峻的挑战，包括主要航道和港口区域的交通密度急剧增加、航行环境日趋复杂等问题。在这一背景下，海上交通事故，特别是船舶碰撞事故，不仅造成巨大的经济损失，还可能引发严重的环境灾难和人员伤亡 [8]。与此同时，以海上自主水面船舶 (MASS) 为代表的航运智能化革命正在兴起，国际海事组织 (IMO) 已将 MASS 的规范化运作提上议程 [19]。MASS 的出现对航行安全和交通效率提出了前所未有的高要求，其安全、可靠的运行高度依赖于对周围动态环境的精准感知和预判能力 [20]。

在所有感知与决策任务中，对周边船舶未来数分钟内的运动轨迹进行准确、可靠的预测，构成了整个智能航运技术体系的基石。精确的多船轨迹预测是实现自主避碰决策、优化航线规划与燃油效率、以及构建高效智能的船舶交通服务 (VTS) 系统不可或缺的前提。一个能够预见未来并理解复杂交互的系统，才能在多变的海洋环境中做出安全、高效的决策。

### 问题定义与现有方法的局限性

多船轨迹预测问题可以形式化地定义为：给定一个海域内  $N$  艘船舶在过去  $T_{\text{obs}}$  个时间步长的状态序列（包括位置、速度、航向等），目标是预测这些船舶在未来  $T_{\text{pred}}$  个时间步长的联合概率分布，并从中采样出多组可能的未来轨迹。该问题本质上是一个多智能体、时空动态 [3] 且具有高度不确定性和多模态特性 [5] 的复杂预测任务。当前解决该方法主要源自两个方向，但都存在根本性的局限。

#### 传统海事方法的局限性

长期以来，海事领域依赖于明确的物理和规则模型进行风险评估。其中，到达最近点时间 (TCPA) 和最近点距离 (DCPA) 是评估碰撞风险的核心指标，被广泛应用于自动雷达标绘仪 (ARPA) 和自动识别系统 (AIS) 中 [7]。然而，这些传统方法存在以下深刻的局限性：

- **线性假设的脆弱性：** TCPA/DCPA 的计算基于一个强假设，即所有船舶都将保持恒速恒向的直线运动。这一假设在船舶进行变速、转向等机动避碰时会立即失效，导致预测结果严重偏离实际情况，无法为决策提供可靠依据。
- **上下文信息的缺失：** 这些指标本质上是两个船舶相对运动向量的几何投影，它们忽略了丰富的上下文信息，如船舶的相对方位、航行态势（如对遇、交叉、追越）等。这可能导致在某些符合《国际海上避碰规则》(COLREGs) 的特定场景下产生风险误判 [8]。
- **低信噪比问题：** 在港口、狭窄水道等交通繁忙水域，大量船舶在物理上彼此靠近，若仅依赖固定的 TCPA/DCPA 阈值进行预警，会产生海量的虚警，严重干扰操作员的判断，降低整个预警系统的可信度和可用性。
- **建模能力的匮乏：** 基于这些指标构建的风险模型，无论是简单的加权求和公式还是基于专家经验的模糊逻辑系统 [9]，都难以从海量历史数据中学习到船舶间复杂、非线性的交互模式，其模型能力和泛化性受到极大限制。

### 通用 AI 模型的领域鸿沟

近年来，基于深度学习的轨迹预测模型，特别是图神经网络 (GNN) [2] 和 Transformer[1]，在自动驾驶车辆和行人轨迹预测等领域取得了突破性进展。然而，将这些模型直接应用于海事领域

会面临一个显著的“领域鸿沟”：

- **交互逻辑的根本差异：**陆基模型通常假设智能体间的交互强度与它们之间的欧几里得距离成反比。但在海上，一艘远在数海里之外但航向与本船交叉的船舶，其威胁等级可能远高于一艘近在咫尺但航向平行的船舶。船舶间的交互本质上是由 COLREGs 定义的、基于未来碰撞风险的逻辑，而非简单的物理邻近性。
- **动力学特性的不同：**船舶具有巨大的惯性和较长的响应时间，其运动学和动力学模型与陆地车辆或行人截然不同。通用模型中对智能体快速加减速和转向的假设在海事领域并不成立。

这种 AI 模型理解的“邻近”与海事安全所关注的“风险”之间的语义鸿沟，是导致现有通用模型在海事领域水土不服的根本原因。模型被迫从数据中艰难地学习这种复杂的、非欧几里得的交互规则，过程不仅数据效率低下，而且容易学到虚假的关联，导致模型鲁棒性差 [12]。

本文提出的解决方案与贡献

为解决上述挑战，本方案提出动态时空图 Transformer (D-STGT) 框架。其核心思想是通过将海事领域的先验安全准则作为一种强归纳偏置，深度嵌入到先进的深度学习模型架构中，从而弥补传统海事方法的模型能力不足与通用 AI 模型的领域知识缺乏之间的鸿沟。

我们的具体方案是，不再将 TCPA/DCPA 作为模型的最终输出或一个独立的后处理模块，而是创新性地将其用作动态构建时空交互图的“边生成器”。在这个图中，节点代表船舶，而连接节点的边及其权重则由两船间的 TCPA 和 DCPA 值动态确定。这样，图的拓扑结构本身就具备了明确的、随时间演变的海事安全意义，它直接编码了场景中的风险关系。

在此基础上，我们设计了一种海事智能体感知图注意力机制。该机制通过差分处理，能够在统一的框架内同时关注两种不同尺度的信息：一是船舶自身的运动惯性，通过对其历史轨迹的自注意力来捕捉；二是由潜在碰撞风险驱动的船舶间交互，通过在风险图上的图注意力来建模。

本研究的主要贡献可概括为以下三点：

1. **提出一种新颖的动态图构建机制：**该机制由海事安全指标 (TCPA/DCPA) 驱动，使图的拓扑结构和边权重能够实时、准确地反映船舶间的碰撞风险等级，而非简单的空间邻近性。这为深度学习模型引入了宝贵的、可解释的领域知识。
2. **设计一种创新的图注意力编码器：**提出了海事智能体感知图注意力 (Maritime Agent-Aware Graph Attention) 机制，首次将代理感知注意力的思想从序列模型迁移并适配到动态图模型中，从而能够有效解耦和建模多尺度的船舶时空动力学。
3. **构建一个完整的、端到端的海事轨迹预测框架：**该框架整合了动态风险图、海事感知注意力编码器和多模态解码器，为解决复杂海事场景下的轨迹预测问题提供了一个全面的、性能优越且具有更强可解释性的解决方案。

相关工作

基于深度学习的轨迹预测

从 RNN 到 Transformer

早期的研究工作广泛采用循环神经网络 (RNN) 及其高级变体，如长短期记忆网络 (LSTM) 和门控循环单元 (GRU)，来捕捉轨迹数据中的时序依赖性。其中，Alahi 等人提出的 Social-LSTM[11]

是一个里程碑式的工作，它通过一个“社交池化层”首次实现了在深度学习框架内对邻近智能体之间的交互进行建模。然而，RNN 类模型固有的序列计算模式限制了其并行处理能力，并且在捕捉长期依赖方面存在梯度消失或爆炸的问题。为了克服这些限制，Vaswani 等人提出的 Transformer 架构被迅速引入轨迹预测领域。Transformer 完全依赖于自注意力机制，能够并行处理序列中的所有元素，并在捕捉长距离依赖关系方面展现出卓越的性能，已成为当前最先进的轨迹预测模型的基础架构 [1]。

### 多模态预测

现实世界中，智能体的未来行为具有内在的不确定性，一个给定的历史轨迹可能对应多个合理且可能发生的未来。因此，单路径预测模型无法充分描述这种不确定性。为了解决这个问题，多模态预测成为了研究的主流。目前主要有两种技术路线：

- **基于条件变分自编码器 (CVAE) 的方法：**这类方法通过一个编码器将未来轨迹信息编码到一个低维的潜在空间中，并在解码时从一个简单的先验分布（通常是标准正态分布）中采样一个潜在变量  $z$ ，与历史信息共同作为解码器的输入，从而生成多样化的未来轨迹 [5]。通过多次采样  $z$ ，可以得到一组不同的预测结果。
- **基于混合密度网络 (MDN) 的方法：**这类方法不直接预测轨迹点，而是让神经网络的输出层预测一个概率混合模型的参数，通常是高斯混合模型 (GMM) 的均值、方差和混合系数 [6]。通过这种方式，模型直接学习了未来位置的概率分布，能够显式地表示多模态特性。

我们的 D-STGT 框架将采用 CVAE 方法来生成多模态预测，因为它在生成高质量、多样化的轨迹样本方面表现出色。

### 时空图神经网络 (ST-GNNs)

#### ST-GNNs 在交通领域的应用

时空图神经网络 (ST-GNNs) 是专门为处理网络化时空数据（如交通网络、传感器网络）而设计的深度学习架构 [3]。它们通过将图神经网络 (GNN) 与序列模型（如 RNN 或时间卷积网络）相结合，能够同时捕捉数据中的空间依赖性（通过图的拓扑结构）和时间动态性（通过节点特征的时间演变）。在智能交通系统 (ITS) 领域，ST-GNNs 已在交通流量预测、路网速度预测等任务上取得了最先进的成果。

#### 动态图神经网络

传统的 ST-GNNs 通常假设图的拓扑结构是静态的，这在许多真实场景中并不成立。例如，在船舶交通中，船舶间的交互关系是随时间和相对位置动态变化的。动态图神经网络 (Dynamic GNNs) 正是为了处理这种时变图结构而提出的 [4]。根据一篇全面的综述 [4]，动态 GNNs 可分为两大类：

- **离散时间动态图 (Discrete-Time Dynamic Graphs, DTDG)：**将动态图视为一系列在离散时间点上的图快照。模型在每个快照上应用 GNN，然后使用序列模型（如 RNN 或 Transformer）来捕捉快照之间的演变规律。
- **连续时间动态图 (Continuous-Time Dynamic Graphs, CTDG)：**将图的演变视为一系列连续时间戳上的事件（如边的增删）。这类模型通常使用时间点过程或专门的 RNN 来处理异步事件。

我们的 D-STGT 模型属于 DTDG 的范畴，在每个时间步，我们都会根据最新的船舶状态 (TCPA/DCPA) 重新计算生成一个新的风险交互图快照。

## 自适应图学习

与我们的工作相关的另一个研究分支是自适应图学习，其核心思想是图的邻接矩阵不再是预先定义的，而是作为模型参数的一部分，从数据中端到端地学习得到。例如，一些工作通过两个可学习的嵌入矩阵的点积来生成自适应邻接矩阵。虽然这种方法具有高度的灵活性，但它缺乏明确的物理或规则约束，可能导致学习到的图结构难以解释，并且在数据稀疏或存在噪声时可能不稳定。相比之下，我们的方法是一种知识引导的图结构学习，我们不让模型从零开始学习图结构，而是利用海事领域百年沉淀的安全准则（TCPA/DCPA）来直接构建图。这种设计旨在将领域知识作为一种强归纳偏置注入模型，以期提高模型的鲁棒性、数据效率和可解释性。

## 智能体交互建模

### AgentFormer 的核心机制

AgentFormer[1] 是近年来在多智能体轨迹预测领域具有重要影响力的工作。其核心思想是通过将多智能体的时空轨迹数据“扁平化”（flattening）为一个单一的长序列，从而能够在统一的 Transformer 框架内，利用自注意力机制联合建模（jointly model）时间和社交两个维度。这种方法打破了传统模型中先进行时间编码再进行社交交互建模的串行模式，允许任意智能体在任意过去时刻的状态直接影响另一智能体在未来任意时刻的状态。

### 代理感知注意力（Agent-Aware Attention）

标准的自注意力机制在处理扁平化序列时，会丢失每个元素所属的智能体身份信息。为了解决这个问题，AgentFormer 提出了其关键创新——代理感知注意力。该机制为智能体内部的交互（intra-agent，即一个智能体自身在不同时间点之间的交互）和智能体之间的交互（inter-agent）学习两套独立的查询（Query）、键（Key）和值（Value）线性投影矩阵。在计算注意力时，对于查询来自智能体  $i$ 、键来自智能体  $j$  的情况，如果  $i = j$ ，则使用内部交互的参数；如果  $i \neq j$ ，则使用外部交互的参数。这种设计巧妙地在统一的注意力计算中保留了智能体的身份信息，使得模型能够差分地学习一个智能体自身的运动学规律和它与其他智能体之间的社交动态。这一思想是我们提出的“海事智能体感知图注意力”机制的直接灵感来源。

### 其他交互建模方法

除了 AgentFormer，研究界还探索了多种交互建模范式。例如，许多工作使用图卷积网络（GCN）或图注意力网络（GAT）[2] 来显式地建模智能体之间的交互关系，其中智能体为节点，它们之间的关系（如距离）为边。最近，还有工作引入了因果推断的思想，如 CRiTIC 模型 [12]，它尝试学习智能体之间的因果关系图，并设计因果注意力门控机制来过滤掉来自非因果相关智能体的信息，以提高模型的鲁棒性。这些工作都凸显了精确建模交互关系的重要性，也为我们选择并改编 AgentFormer 的先进机制提供了有力的背景支撑。

## 海事碰撞风险评估

### 传统指标与模型

海事领域的碰撞风险评估长期以来都围绕着 TCPA 和 DCPA 这两个核心指标展开。它们是 ARPA 和 AIS 系统的基本功能，为船员提供了关于会遇态势最直观的量化信息 [7]。早期的一些风险指数模型尝试将这两个指标通过加权公式结合起来，但这些模型存在根本性缺陷，例如将不同物理量纲的变量（距离和时间）直接相加，以及权重系数仅凭经验设定，缺乏自适应性。

### 船舶领域模型（Ship Domain）

为了克服 TCPA/DCPA 的局限性，船舶领域模型被提出。它定义了船舶周围一个需要保持无其他障碍物侵入的安全区域，其形状和大小通常与船速、船长、交通密度等因素有关 [8]。基于领域



模型，碰撞风险可以被定义为领域被侵犯的程度或概率。尽管领域模型比单一的 TCPA/DCPA 指标更为全面，但它们大多仍是基于几何规则或专家系统驱动的，难以与现代深度学习模型进行端到端的无缝集成。

#### 研究空白

纵观相关工作，可以发现现代 AI 轨迹预测技术与传统海事风险评估方法之间存在一条明显的鸿沟。前者拥有强大的数据驱动学习能力但缺乏领域知识，后者拥有坚实的领域知识但模型表达能力有限。本研究工作的核心定位正是要填补这一空白。我们不将海事风险评估视为一个独立的预处理或后处理步骤，而是将其核心逻辑（即 TCPA/DCPA 所蕴含的风险信息）作为构建深度学习模型架构的基本元素，从而实现领域知识与数据驱动方法的深度、有机融合。这种融合体现了一种重要的研究范式转变：从单纯依赖数据学习交互，转向利用领域知识指导交互的学习。这一转变不仅是技术上的创新，也反映了对复杂系统建模更深层次的理解——即一个高效的模型应当是数据驱动与知识引导的有机结合体。

AgentFormer 中对智能体内部和外部动态的解耦思想，与 ST-GNN 中对时间动态和空间依赖的分别处理，在概念上不谋而合，都指向了对复杂系统中不同尺度作用进行分解建模的共同原则。这进一步验证了我们提出的“海事智能体感知图注意力”机制的合理性，它并非简单的思想拼接，而是一种在图结构上对这一通用建模原则的自然延伸和实现，旨在将 AgentFormer 的灵活性与 GNN 的结构化归纳偏置相结合，以应对海事领域的独特挑战。

#### 方法论：动态时空图 Transformer

本部分将详细阐述我们提出的动态时空图 Transformer (D-STGT) 模型的架构设计和数学原理。

Table 1: 符号与定义

符号	定义
$N$	场景中的船舶数量
$T_{\text{obs}}$	历史轨迹观测时长（时间步数）
$T_{\text{pred}}$	未来轨迹预测时长（时间步数）
$s_t^i$	船舶 $i$ 在时间步 $t$ 的状态向量, e.g., $s_t^i = (x_t^i, y_t^i, v_x^i, v_y^i)$
$X_i$	船舶 $i$ 的观测轨迹序列, $X_i = (s_1^i, \dots, s_{T_{\text{obs}}}^i)$
$Y_i$	船舶 $i$ 的真实未来轨迹序列, $Y_i = (s_{T_{\text{obs}}+1}^i, \dots, s_{T_{\text{obs}}+T_{\text{pred}}}^i)$
$\hat{Y}_i^k$	模型对船舶 $i$ 的第 $k$ 个多模态预测未来轨迹
$G_t$	在时间步 $t$ 的动态海事安全图, $G_t = (V_t, E_t, W_t)$
$V_t$	图 $G_t$ 的节点集合，代表所有船舶
$E_t$	图 $G_t$ 的边集合，代表船舶间的风险关系
$A_t$	图 $G_t$ 的加权邻接矩阵
$h_t^i$	船舶 $i$ 在时间步 $t$ 的节点特征向量（嵌入表示）
$H^{(l)}$	模型第 $l$ 层的节点特征矩阵
$W_Q, W_K, W_V$	Transformer 注意力机制中的查询、键、值投影矩阵
$z$	CVAE 中的潜在变量
$\mathcal{L}$	模型的总损失函数

## 问题形式化与框架概览

给定一个包含  $N$  艘船舶的场景,我们的任务是基于它们在过去  $T_{\text{obs}}$  个时间步的观测轨迹  $\{X_1, X_2, \dots, X_N\}$ , 其中  $X_i \in \mathbb{R}^{T_{\text{obs}} \times D_{\text{in}}}$  且  $D_{\text{in}}$  是输入状态维度, 来预测它们在未来  $T_{\text{pred}}$  个时间步的联合轨迹分布。具体来说, 模型需要学习一个映射函数  $f: \{X_i\}_{i=1}^N \rightarrow P(\{\hat{Y}_i\}_{i=1}^N)$ , 其中  $P$  是未来联合轨迹的概率分布,  $\hat{Y}_i \in \mathbb{R}^{T_{\text{pred}} \times D_{\text{out}}}$ 。由于未来具有多模态性, 我们期望能从该分布中采样出  $K$  组不同的、符合现实的联合轨迹场景  $\{\{\hat{Y}_1^k, \dots, \hat{Y}_N^k\}\}_{k=1}^K$ 。

D-STGT 框架的整体架构如 (图) 所示, 它主要由三个核心模块构成:

- (1) **动态海事安全图构建模块**: 该模块在每个时间步接收所有船舶的实时状态, 计算两两之间的 TCPA 和 DCPA, 并据此生成一个动态加权图, 其拓扑结构显式地编码了场景中的碰撞风险。
- (2) **海事智能体感知图注意力编码器**: 该模块是模型的核心, 它以动态图序列和船舶历史轨迹为输入, 通过我们设计的海事智能体感知图注意力机制, 学习每艘船在时空交互下的动态表征。
- (3) **多模态轨迹解码器**: 该模块基于 CVAE 框架, 将编码器输出的上下文表征与从先验分布中采样的潜在变量相结合, 通过一个自回归解码器生成多样化的未来轨迹预测。

## 动态海事安全图构建

这一模块的目标是将海事领域的先验安全知识转化为深度学习模型可以直接利用的结构化信息。我们摒弃了传统 GNN 中基于固定距离或全连接的图构建方式, 转而采用一种由风险驱动的动态图构建策略。

### 节点定义:

在任意时间步  $t$ , 图  $G_t = (V_t, E_t, A_t)$  的节点集合  $V_t$  由场景中的  $N$  艘船舶构成。每艘船  $i$  的状态向量  $s_t^i$  (包含经纬度、速度分量等) 首先通过一个共享的 MLP 网络  $f_{\text{embed}}$  得到一个高维特征空间, 得到该节点的初始特征  $h_t^i = f_{\text{embed}}(s_t^i)$ 。

### 动态图权重定义:

这是本方法的关键创新。在每个时间步  $t$ , 任意两艘船  $i$  和  $j$  之间的边的存在性及其权重, 完全由它们此刻的会遇态势决定, 具体通过 TCPA 和 DCPA 来量化。

### TCPA/DCPA 计算:

给定两船的状态  $s_t^i = (x_t^i, y_t^i, v_x^i, v_y^i)$  和  $s_t^j = (x_t^j, y_t^j, v_x^j, v_y^j)$ , 我们首先计算它们的相对位置向量  $\vec{P}_{\text{rel}} = (x_t^j - x_t^i, y_t^j - y_t^i)$  和相对速度向量  $\vec{V}_{\text{rel}} = (v_x^j - v_x^i, v_y^j - v_y^i)$ 。根据运动学原理, DCPA 和 TCPA 的计算公式如下:

$$\text{DCPA}_{ij} = \frac{|(x_t^j - x_t^i)(v_y^j - v_y^i) - (y_t^j - y_t^i)(v_x^j - v_x^i)|}{\sqrt{(v_x^j - v_x^i)^2 + (v_y^j - v_y^i)^2}} = \frac{|\vec{P}_{\text{rel}} \times \vec{V}_{\text{rel}}|}{|\vec{V}_{\text{rel}}|} \quad (1)$$

$$\text{TCPA}_{ij} = -\frac{(x_t^j - x_t^i)(v_x^j - v_x^i) + (y_t^j - y_t^i)(v_y^j - v_y^i)}{(v_x^j - v_x^i)^2 + (v_y^j - v_y^i)^2} = -\frac{\vec{P}_{\text{rel}} \cdot \vec{V}_{\text{rel}}}{|\vec{V}_{\text{rel}}|^2} \quad (2)$$

其中,  $\times$  表示二维向量的叉乘大小,  $\cdot$  表示点积。

### 边生成与权重计算:

我们引入海事实践中常用的安全阈值, 如最小安全 DCPA  $D_{\text{safe}}$  和最大预警 TCPA  $T_{\text{safe}}$ [9]。一条边  $e_{ij,t}$  在图  $G_t$  中存在, 当且仅当两船构成潜在的碰撞威胁, 即  $0 < \text{TCPA}_{ij} \leq T_{\text{safe}}$  且  $\text{DCPA}_{ij} \leq D_{\text{safe}}$ 。这确保了图的连接只在具有实际避碰意义的船舶对之间建立, 形成了一个稀疏且语义明确

的风险图。边的权重  $w_{ij,t}$  被设计为风险程度的函数，使得风险越高的交互获得越大的权重。一个简单而有效的实现是使用指数衰减函数：

$$w_{ij,t} = \exp \left( -\alpha \frac{\text{DCPA}_{ij}}{D_{\text{safe}}} - \beta \frac{\text{TCPA}_{ij}}{T_{\text{safe}}} \right)$$

其中  $\alpha$  和  $\beta$  是可调的超参数。这样，当 DCPA 和 TCPA 趋近于 0 时，权重趋近于 1（最高风险）；当它们接近阈值时，权重减小。所有这些权重构成了时间步  $t$  的动态加权邻接矩阵  $A_t$ 。

通过这种方式，我们构建了一个动态图序列  $\{G_1, G_2, \dots, G_{T_{\text{obs}}}\}$ ，它将作为我们编码器模块的输入。这种知识注入的设计，使得模型不必从零开始学习何为“危险”，而是直接在一个编码了“危险”的结构上进行推理，极大地降低了学习难度并提升了模型的可解释性。

海事智能体感知图注意力编码器

为了有效处理船舶运动中自身惯性与外部交互的耦合影响，我们设计了海事智能体感知图注意力（Maritime Agent-Aware Graph Attention, MA-GAT）编码器。其核心思想是为每个节点（船舶）的特征更新提供两条并行的信息流：一条处理其自身的时间动态，另一条处理其在动态风险图上的空间交互。

机制设计：

编码器由  $L$  个 MA-GAT 层堆叠而成。在第  $l$  层，对于每个节点  $i$ ，其特征向量  $h_i^{(l)}$  的更新  $h_i^{(l+1)}$  由以下两部分融合而成：

- **内部时序注意力 (Intra-Ship Temporal Attention):** 此部分旨在捕捉单艘船舶自身的运动模式和惯性。我们对船舶  $i$  在所有历史观测时间步的特征序列  $\{h_{1,i}^{(l)}, \dots, h_{T_{\text{obs}},i}^{(l)}\}$  应用一个标准的多头自注意力（Multi-Head Self-Attention, MHSA）机制，允许模型在更新当前时刻的特征时，动态地关注历史轨迹中的关键节点（例如，转向的起始点）。

$$h_{\text{intra},i}^{(l)} = \text{MHSA}_{\text{intra}} \left( q = h_{T_{\text{obs}},i}^{(l)}, K = V = \{h_{t,i}^{(l)}\}_{t=1}^{T_{\text{obs}}} \right) \quad (4)$$

其中，查询向量  $q$  是当前时刻的特征，而键  $K$  和值  $V$  是历史特征序列。 $\text{MHSA}_{\text{intra}}$  使用独立的参数集  $(W_Q^{\text{intra}}, W_K^{\text{intra}}, W_V^{\text{intra}})$ 。

- **外部风险交互注意力 (Inter-Ship Risk-Interactive Attention):** 此部分用于建模船舶之间基于碰撞风险的交互。我们采用图注意力网络（GAT）[2] 的思想，但在动态海事安全图  $G_{T_{\text{obs}}}$  上进行。对于节点  $i$ ，它会关注其所有邻居  $j \in \mathcal{N}_i$ （即在  $T_{\text{obs}}$  时刻与其构成碰撞风险的船舶）。注意力系数  $\alpha_{ij}^{(l)}$  不仅取决于节点特征，还受到我们之前计算的风险权重  $w_{ij,T_{\text{obs}}}$  的调制。注意力系数计算如下：

$$e_{ij}^{(l)} = \text{LeakyReLU} \left( \bar{a}^{(l)T} [h_i^{(l)} W_Q^{\text{inter}} || h_j^{(l)} W_K^{\text{inter}}] \right) \cdot w_{ij,T_{\text{obs}}} \quad (5)$$

$$\alpha_{ij}^{(l)} = \text{softmax}_j(e_{ij}^{(l)}) = \frac{\exp(e_{ij}^{(l)})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik}^{(l)})} \quad (6)$$

其中  $||$  表示拼接， $\bar{a}^{(l)}$  是一个可学习的权重向量， $W_Q^{\text{inter}}, W_K^{\text{inter}}$  是另一套独立的投影矩阵。风险权重  $w_{ij,T_{\text{obs}}}$  作为调制因子，增强了高风险邻居的影响力。聚合后的特征为：

$$h_{\text{inter},i}^{(l)} = \sigma \left( \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(l)} (h_j^{(l)} W_V^{\text{inter}}) \right) \quad (7)$$

其中  $\sigma$  是非线性激活函数（如 ELU）。同样，这里也使用多头机制来稳定学习过程。

特征融合与层级更新:

在得到  $h_{\text{intra},i}^{(l)}$  和  $h_{\text{inter},i}^{(l)}$  后, 我们将它们融合以形成该层的最终输出。一个简单的融合方式是相加后接层归一化 (LayerNorm) 和前馈网络 (FFN):

$$h_i^{(l+1)} = \text{LayerNorm} \left( h_i^{(l)} + \text{FFN} \left( \text{LayerNorm} \left( h_{\text{intra},i}^{(l)} + h_{\text{inter},i}^{(l)} \right) \right) \right) \quad (8)$$

通过堆叠  $L$  个这样的 MA-GAT 层, 模型可以学习到深层次、多尺度的时空交互表征  $h_i^{(L)}$ , 该表征既包含了单船的运动意图, 也融入了对周围环境风险的响应。

计算复杂度分析:

该机制的计算复杂度主要由两部分构成。内部时序注意力的复杂度为  $O(N \cdot T_{\text{obs}}^2 \cdot D)$ , 其中  $D$  是特征维度。外部风险交互注意力的复杂度为  $O(N \cdot D^2 + |E_{T_{\text{obs}}}| \cdot D)$ , 其中  $|E_{T_{\text{obs}}}|$  是风险图中的边数。由于海事风险图通常是稀疏的 ( $|E_{T_{\text{obs}}}| \ll N^2$ ), 且  $T_{\text{obs}}$  通常是固定的、不太大的值, 因此该架构的计算开销在实际应用中是可控的。

多模态轨迹解码器与训练目标

为了捕捉未来轨迹的不确定性, 我们采用基于 CVAE 的解码器框架 [5]。

CVAE 解码器:

在训练阶段, 我们使用一个额外的编码器 (后验网络) 将真实的未来轨迹  $Y$  和编码器输出的上下文表征  $H^{(L)}$  一起编码为后验分布  $q(z|Y, H^{(L)})$  的参数 (均值  $\mu_z$  和方差  $\sigma_z^2$ )。在推理阶段, 我们从一个简单的先验分布  $p(z) = \mathcal{N}(0, I)$  中采样一个潜在变量  $z$ 。然后, 将上下文表征  $h_i^{(L)}$  和潜在变量  $z_i$  (每个智能体可以有独立的  $z$ ) 拼接后, 输入到一个基于 GRU 的自回归解码器中。解码器在每个时间步  $t$  预测下一位置的分布, 并将预测结果作为下一时间步的输入, 直到生成完整的未来轨迹  $\hat{Y}_i$ 。

训练目标:

模型的训练目标是最大化证据下界 (Evidence Lower Bound, ELBO), 这等价于最小化一个复合损失函数  $\mathcal{L}$ :

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda_{\text{KL}} \mathcal{L}_{\text{KL}} \quad (9)$$

重构损失 ( $\mathcal{L}_{\text{recon}}$ ):

该损失项旨在使预测轨迹尽可能接近真实轨迹。为了处理多模态输出, 我们采用“最优样本”策略, 即在  $K$  个采样预测中, 只惩罚与真实轨迹最接近的那个。我们使用 Huber 损失代替 L2 损失, 以增强模型对异常值 (如 AIS 数据中的瞬时跳点) 的鲁棒性。

$$\mathcal{L}_{\text{recon}} = \frac{1}{N \cdot T_{\text{pred}}} \sum_{i=1}^N \min_{k \in \{1, \dots, K\}} \sum_{t=T_{\text{obs}}+1}^{T_{\text{obs}}+T_{\text{pred}}} \text{Huber}(\hat{y}_{k,t}^i - y_t^i) \quad (10)$$

KL 散度损失 ( $\mathcal{L}_{\text{KL}}$ ):

这是 CVAE 框架的正则化项, 用于约束后验分布  $q(z|Y, H^{(L)})$  接近于先验分布  $p(z)$ , 防止模型过拟合到训练数据, 并确保在推理时从先验分布中采样是有效的。

$$\mathcal{L}_{\text{KL}} = \frac{1}{N} \sum_{i=1}^N D_{\text{KL}} \left( q(z_i|Y_i, h_i^{(L)}) \| p(z_i) \right) \quad (11)$$

$\lambda_{\text{KL}}$  是一个平衡两个损失项的超参数。通过优化这个总损失, D-STGT 模型能够学习生成既准确又多样化的未来轨迹。

该方法论的核心在于，它并非一个纯粹的数据驱动“黑箱”，而是一个“知识注入”的模型。传统海事安全规则（TCPA/DCPA）不再是与模型无关的外部约束，而是被转化为模型内部的结构性偏置。这种设计使得模型在学习过程中，其假设空间被有效地约束在符合海事安全逻辑的范围内。这预计将带来多重优势：首先是更高的样本效率，因为模型无需从海量数据中“重新发现”基本的避碰原则；其次是更强的鲁棒性，因为其交互逻辑基于普适的安全规则，不易被训练数据中的虚假相关性所干扰，在面对分布外（out-of-distribution）的罕见但关键的会遇场景时，表现可能更稳定；最后是更好的泛化能力，因为安全原则是跨地域、跨场景通用的，有助于模型从一个海域学习到的知识迁移到另一个海域。

### 实验设计

为了全面、严谨地评估我们提出的 D-STGT 模型的性能，我们设计了以下实验方案，包括数据集选择、评估指标、基线模型对比以及消融研究。

### 数据集与预处理

#### 数据源：

本研究将主要采用公开的大规模 AIS 数据集进行训练和评估。首选数据源是美国国家海洋和大气管理局（NOAA）的 Marine Cadastre 项目 [16]。该项目提供了自 2009 年以来美国沿海水域的 AIS 数据，数据经过初步处理，具有一分钟的采样率，并以每日或每月的形式提供，覆盖范围广，时间跨度长。此外，我们还将探索其他公开的 AIS 数据源，以增加数据的多样性，测试模型的泛化能力。

#### 数据挑战与预处理：

原始 AIS 数据并非为科研设计，存在诸多固有问题，必须进行细致的预处理才能使用 [17]。我们的预处理流程将包括以下关键步骤：

- **数据清洗与插值：** AIS 数据常因 VHF 信号传播问题、信道拥塞等原因出现缺失值、异常跳点和噪声。我们将采用移动窗口平滑滤波来抑制高频噪声，并使用线性或更高阶的插值方法（如样条插值）来填补短时间（例如几分钟内）的信号丢失，以重建连续的轨迹。
- **处理不一致性：** 静态和航次相关信息（如船型、目的地、吃水、航行状态）依赖船员手动输入，常存在错误或未及时更新的情况。我们将设计规则和启发式算法来校验这些信息，例如，通过分析轨迹模式来推断真实的航行状态（如“在航”、“锚泊”或“靠泊”）。
- **异常行为识别：** AIS 系统存在被恶意利用的风险，如 AIS 欺骗（spoofing，发送虚假位置信息）和“暗黑航运”（dark shipping，故意关闭 AIS 发射器）[18]。我们将实现一个异常检测模块，基于运动学不可能性（如瞬时速度过高）和历史航行模式偏离等特征，来识别并过滤掉这些不可靠的轨迹数据，以保证训练数据的质量和安全性。
- **轨迹切片与场景构建：** 将清洗和重建后的连续轨迹数据流，按照固定的时间窗口（例如，观测 8 分钟，预测 12 分钟）进行切片，生成大量的轨迹片段。然后，将同一时间窗口内地理位置相近的船舶轨迹片段组合成一个“场景”，作为模型的一个输入样本。

### 评估指标

评估轨迹预测模型需要一个多维度的指标体系，单纯的精度指标不足以全面反映模型性能，尤其是在安全攸关的海事领域。我们的评估将采用以下指标：

标准位移误差：

- **最小平均位移误差 (minADE<sub>K</sub>):** 对于 K 个多模态预测样本，计算每个智能体预测轨迹与真实轨迹在所有时间步上的平均 L2 距离，并取 K 个样本中的最小值。然后对场景中所有智能体求平均。这是衡量预测中心线准确性的标准指标 [11]。

$$\text{minADE}_K = \frac{1}{N} \sum_{i=1}^N \min_{k \in \{1..K\}} \frac{1}{T_{\text{pred}}} \sum_{t=T_{\text{obs}}+1}^{T_{\text{obs}}+T_{\text{pred}}} \|\hat{y}_{k,t}^i - y_t^i\|_2$$

- **最小最终位移误差 (minFDE<sub>K</sub>):** 与 minADE 类似，但只计算预测终点与真实终点之间的 L2 距离。它衡量了模型对智能体最终意图的预测能力 [11]。

$$\text{minFDE}_K = \frac{1}{N} \sum_{i=1}^N \min_{k \in \{1..K\}} \|\hat{y}_{k,T_{\text{obs}}+T_{\text{pred}}}^i - y_{T_{\text{obs}}+T_{\text{pred}}}^i\|_2$$

交互感知联合指标：

最近的研究指出，minADE/FDE 这类边际指标存在严重缺陷，因为它们允许模型通过从不同的预测样本中为不同的智能体“挑选”最优轨迹来获得高分，而这些最优轨迹在同一个样本中可能并不共存，甚至可能相互冲突（如碰撞）[10]。为了评估模型对整个场景联合动态的建模能力，我们引入：

- **联合平均/最终位移误差 (JADE/JFDE):** 这类指标首先计算每个预测样本内所有智能体的平均误差，然后再在 K 个样本中取最小值。这要求模型在至少一个完整的预测场景中对所有智能体都做出合理的预测 [10]。
- **碰撞率 (Collision Rate - CR):** 这是评估模型安全性的最直接指标。我们将定义一个基于船舶尺寸的碰撞阈值。CR 被计算为在所有测试场景中，模型预测的 K 个未来样本中至少包含一对碰撞轨迹的样本所占的比例 [10]。一个理想的模型应该在保证高精度的同时，具有极低的碰撞率。

基线模型

为了证明 D-STGT 的优越性，我们将与以下几类代表性的基线模型进行公平对比：

- **经典模型:** Social-LSTM[11]，作为基于 RNN 和社交池化的开创性工作，代表了早期交互建模方法的水平。
- **基于 GNN 的模型:** STGAT，一个代表性的时空图注意力网络，用于对比我们的图注意力机制。Trajectron++，一个强大且模块化的图结构化 RNN 模型，在多个基准测试中表现出色。
- **基于 Transformer 的模型:** AgentFormer[1]，这是我们模型思想的核心来源之一，也是当前最先进的轨迹预测模型之一。与 AgentFormer 的直接对比是验证我们创新的关键。
- **领域自适应变体:** 为了更公平地隔离和评估我们提出的海事智能体感知注意力机制的贡献，我们将为 AgentFormer 和 STGAT 等通用模型创建“海事变体”。具体做法是，用我们提出的 TCPA/DCPA 驱动的动态图来替换它们原有的基于欧几里得距离的交互模块。通过比较“AgentFormer + 我们的图”与“D-STGT”，我们可以更清晰地看到我们设计的注意力机制相较于标准 Transformer 注意力的优势。

## 消融研究

为了系统性地验证我们模型中每个创新组件的必要性和有效性，我们将进行一系列详尽的消融实验。这对于理解模型行为、证明设计合理性至关重要。具体的消融研究设计如表2所示。

Table 2: 消融研究设计

模型变体	被消融/替换的组件	待验证的假设
D-STGT w/o Dynamic Graph	使用基于距离的静态图（在 $t=1$ 时构建一次）替换动态图生成。	动态更新的图拓扑对于捕捉时变的海事风险至关重要，静态图无法适应交互关系的变化。
D-STGT w/ Proximity Edges	边的生成和权重基于船舶间的欧几里得距离，而非 TCPA/DCPA。	由海事安全指标驱动的图边比简单的物理邻近性更能有效、准确地建模避碰交互。
D-STGT w/ Standard GAT	使用标准的 GAT 层替换海事智能体感知图注意力层，即不对内部时序和外部交互进行解耦。	解耦内部时序惯性和外部风险交互的注意力机制能更精确地学习船舶动力学，避免信息混淆。
D-STGT w/o Intra-Attention	在 MA-GAT 层中，仅保留外部风险交互注意力分支，移除内部时序自注意力分支。	明确地建模船舶自身的运动惯性对于准确预测至关重要，仅靠交互信息不足。
D-STGT w/o Inter-Attention	在 MA-GAT 层中，仅保留内部时序自注意力分支，移除外部风险交互注意力分支。	明确地建模船舶间的风险交互是多智能体预测的核心，缺少该模块将导致模型无法处理避碰场景。

通过这一系列精心设计的实验，我们不仅能够量化 D-STGT 相对于现有技术的性能提升，还能深入洞察其内部各组件的贡献，为模型的有效性提供强有力的实证支持。

### 预期影响与伦理考量

本研究不仅旨在提出一个性能更优的预测模型，更期望通过引入新的建模范式，对智能航运领域产生深远影响，并审慎地探讨其带来的伦理挑战。

### 可解释性、因果推断与信任

在安全攸关的系统中，模型的准确性固然重要，但其决策过程的透明度和可信度同样关键。一个无法被人类理解和信任的“黑箱”模型，在实际应用中将面临巨大的障碍。

### 内在可解释性：

D-STGT 的设计本身就追求更高的可解释性。与那些从数据中学习隐式交互的通用模型不同，我们的动态海事安全图提供了一个清晰的窗口来窥探模型的“注意力”。图中的边直接对应于由 TCPA/DCPA 定义的、人类操作员可以理解的碰撞风险。因此，当分析模型的图注意力权重时，我们可以明确地回答“模型认为哪艘船对哪艘船构成了威胁，以及威胁的程度如何”，这为模型的决策提供了直观的、符合领域逻辑的解释。



与因果推断的结合：

展望未来，本框架为与更前沿的因果推断技术相结合提供了理想的平台，这将进一步增强模型的可解释性、鲁棒性和可信度。

- **因果发现 (Causal Discovery)**：通过分析在不同场景下学习到的图注意力权重的时间序列，我们可以探索船舶交互行为中的因果关系 [12]。例如，我们可以识别出在一次交叉相遇中，是哪艘“让路船” (give-way vessel) 的机动行为“导致”了“直航船” (stand-on vessel) 的轨迹调整，而不仅仅是发现它们之间存在相关性。这对于理解复杂的交通流动态和事故成因分析具有重要意义。
- **反事实推理 (Counterfactual Reasoning)**：我们的模型可以作为一个强大的工具，用于进行反事实分析 [13]。例如，在一次已发生的近距离碰撞事件中，我们可以通过在模型中设定反事实条件（“如果 B 船在关键时刻没有向右转向，而是保持原航向，会发生什么？”），来模拟和推演不同的结局。这种“事后诸葛亮”的能力对于事故复盘、责任界定以及制定更优的避碰策略具有不可估量的价值。
- **建立信任 (Building Trust)**：一个能够提供因果解释（“A 船转向是因为 B 船造成了碰撞风险”）而不仅仅是相关性预测（“A 船和 B 船的轨迹高度相关”）的模型，更容易获得人类操作员（如 VTS 官员或远程操作员）的信任。这种信任是实现高效、安全的人机协同的关键。

鲁棒性与泛化能力

应对分布外 (Out-of-Distribution, OOD) 场景：

深度学习模型的一大挑战是在面对与训练数据分布不同的新场景时的性能下降问题 [14]。海事环境的复杂性（如罕见的会遇态势、异常的天气海况、新的港口布局）使得 OOD 问题尤为突出。我们认为，D-STGT 通过将普适的物理和规则知识 (TCPA/DCPA) 硬编码到其图结构中，相比于纯数据驱动的方法，具有更强的内在鲁棒性。因为无论场景如何变化，这些基本的安全准则始终适用。

领域自适应与元学习：

为了进一步提升模型在不同地理区域和交通模式下的泛化能力，我们规划了未来的研究方向，即引入领域自适应 (Domain Adaptation) 和元学习 (Meta-Learning) 技术。

- **领域自适应**：我们可以将在一个海域（如美国西海岸，源域）训练好的模型，通过无监督或半监督的领域自适应技术，使其适应另一个完全不同的海域（如新加坡海峡，目标域）的交通数据，而无需在目标域进行大量的重新标注和训练。
- **元学习**：我们可以采用“学习如何学习”的元学习框架 [15]。通过在多个不同的源域（多个港口）上进行元训练，模型可以学习到一种通用的、能够快速适应新环境的初始化参数或学习策略。当部署到一个全新的、前所未见的海域时，该模型仅需少量数据就能快速微调，达到理想的性能。

伦理考量

将人工智能深度整合到航运这一关键基础设施中，必须审慎对待其可能带来的伦理和社会挑战。



责任与问责 (Responsibility and Accountability):

当一艘由 AI 系统引导或完全自主的船舶发生事故时, 责任的归属将变得异常复杂。责任方是船东、船舶运营商、AI 软件开发者, 还是数据提供商? [20]。虽然我们的工作通过提升模型的可解释性, 能够为事后技术分析和责任链条的追溯提供重要线索, 但这并不能替代一个清晰、完善的法律和监管框架。本研究的成果应被视为完善未来法规的技术基础之一。

人机协同 (Human-Machine Collaboration):

我们明确将 D-STGT 定位为一个增强人类能力的决策支持工具, 而非意图完全取代人类操作员。其核心价值在于, 将人类从繁重、重复的数据监控中解放出来, 自动识别和高亮显示潜在风险, 从而使人类专家能够将他们的认知资源集中在最高阶的、最复杂的决策上。确保人类在关键决策回路中始终拥有最终的监督和干预权, 是保障系统安全和伦理合规的根本原则。

数据安全与网络安全 (Data and Cybersecurity):

AI 系统的性能高度依赖于其输入数据的质量和完整性。AIS 系统作为一个开放的广播系统, 已被证明易受欺骗、劫持和干扰等网络攻击 [18]。如果我们的模型被输入了被恶意篡改的数据, 可能会做出灾难性的错误预测。因此, 任何基于本研究成果的实际部署, 都必须建立在一个强大的、端到端的数据验证和网络安全框架之上, 确保输入数据的真实性和系统的抗攻击能力。

对劳动力的影响 (Impact on Labor):

航运自动化和智能化不可避免地会对传统海员等岗位的就业产生冲击。这是一个严肃的社会经济问题。我们认为, 负责任的技术发展必须与社会政策相配套。这包括为现有海事从业人员提供系统的再培训和技能升级计划, 使他们能够从传统的直接操作者, 转变为能够与高级智能系统协同工作、进行监督和维护的新型海事专家。

结论

本研究方案提出了一种名为“动态时空图 Transformer (D-STGT)”的新型深度学习框架, 旨在解决复杂海域中多尺度、多智能体船舶轨迹预测的核心挑战。面对传统海事模型表达能力不足和通用 AI 模型领域知识缺乏的困境, 本方案的核心思想是通过深度融合领域知识与先进的 AI 技术, 构建一个既强大又可信的预测模型。

我们重申本方案的两大核心创新: 第一, 提出了一种由海事安全通用准则 TCPA 和 DCPA 驱动的动态图构建方法, 使图的拓扑结构能够实时、显式地编码船舶间的碰撞风险, 为模型注入了宝贵的、可解释的归纳偏置。第二, 设计了一种新颖的海事智能体感知图注意力机制, 通过解耦船舶的自身运动惯性和外部交互影响, 实现了对多尺度船舶动力学的精细化建模。

通过详尽的实验设计, 包括在真实大规模 AIS 数据集上的验证、与多种基线模型的对比以及系统性的消融研究, 我们期望证明 D-STGT 在预测精度、交互真实性和安全性 (以低碰撞率衡量) 方面均能达到或超越当前最先进水平。此外, 本研究对模型的可解释性、鲁棒性以及相关的伦理问题进行了深入探讨, 强调了其作为决策支持工具在人机协同环境下的应用前景。

综上所述, D-STGT 不仅是一个技术上的创新, 更代表了一种将领域知识与数据驱动方法相结合的建模新范式。我们相信, 这项研究的成功将为下一代智能船舶交通管理系统和海上自主水面船舶 (MASS) 的研发提供关键的感知与预测能力, 最终为构建一个更安全、高效和智能的全球海运网络贡献坚实的科学基础和技术支撑。

## References

- [1] R. Yuan, X. Weng, W. Yuan, and K. Kitani, “AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9592–9602.
- [2] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph Attention Networks,” in *International Conference on Learning Representations (ICLR)*, 2018.
- [3] W. Jiang and J. Luo, “Spatio-Temporal Graph Neural Networks: A Survey,” *arXiv preprint arXiv:2203.09773*, 2022.
- [4] S. Kazemi, R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth, and P. Poupart, “A survey of dynamic graph neural networks,” *arXiv preprint arXiv:2404.18211*, 2024.
- [5] C. Shi, “Multimodal Autonomous Vehicle Trajectory Prediction,” Ph.D. dissertation, University of Toronto, 2021.
- [6] C. M. Bishop, *Mixture Density Networks*. Aston University, Neural Computing Research Group Report, 1994.
- [7] J. Zhang, C. Zhang, Z. Liu, H. Yan, and J. Wang, “A Novel Method for Holistic Collision Risk Assessment in the Precautionary Area Using AIS Data,” *Journal of Marine Science and Engineering*, vol. 10, no. 5, p. 649, 2022.
- [8] Z. Liu, M. A. F. G. van der, A. N. D. L. V. L. H. Woltjer, and M. A. D. M. van der, “A review of ship collision risk assessment, hotspot detection and path planning for maritime traffic control in restricted waters,” *The Journal of Navigation*, vol. 74, no. 5, pp. 921–946, 2021.
- [9] A. F. Hisyam, S. Anam, and M. H. Purnomo, “Fuzzy Inference System for Determining Collision Risk of Ship in Madura Strait Using Automatic Identification System,” *International Journal of Fuzzy System Applications*, vol. 7, no. 1, pp. 31–47, 2018.
- [10] T. Gu, J. Chen, B. Wang, and P. Wang, “Joint Metrics Matter: A Better Standard for Trajectory Forecasting,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 24046–24055.
- [11] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social LSTM: Human Trajectory Prediction in Crowded Spaces,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 961–971.
- [12] Z. Li, Z. Wang, Z. Ren, H. Zhang, and J. M. F. Moura, “Causal Attention Gating for Robust Trajectory Prediction in Autonomous Driving,” *arXiv preprint arXiv:2310.13849*, 2023.
- [13] Y. Li, W. Zhan, C. Shi, and M. Tomizuka, “Human Trajectory Prediction via Counterfactual Analysis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 13203–13212.

- [14] J. You, J. Liu, R. Ying, and J. Leskovec, “Graph Learning under Distribution Shifts: A Comprehensive Survey on Domain Adaptation, Out-of-distribution, and Continual Learning,” *arXiv preprint arXiv:2306.12643*, 2023.
- [15] J. Park, S. Park, S. H. Lee, and J. Y. Choi, “MetaTra: Meta-Learning for Generalized Trajectory Prediction in Unseen Domain,” *arXiv preprint arXiv:2303.10052*, 2023.
- [16] National Oceanic and Atmospheric Administration (NOAA), “AIS,” *Data.gov*, accessed May 20, 2024. [Online]. Available: <https://catalog.data.gov/dataset/ais>
- [17] A. Polemis, “AIS and the main categories of AIS challenges,” *Maritime Optima*, 2021. [Online]. Available: <https://www.maritimeoptima.com/blog/ais-challenges>
- [18] E. G. Aarsæther, S. S. Schia, and E. F. Breivik, “AIS Data Vulnerability Indicated by a Spoofing Case-Study,” *Sensors*, vol. 22, no. 17, p. 6428, 2022.
- [19] G. Li, K. Zhang, S. Zhang, and R. Sutton, “Optimizing Multi-Vessel Collision Avoidance Decision Making for Autonomous Surface Vessels: A COLREGs-Compliant Deep Reinforcement Learning Approach,” *Applied Sciences*, vol. 12, no. 1, p. 488, 2022.
- [20] T. Dove, “Maritime AI Ethics,” *Prism → Sustainability Directory*, accessed May 20, 2024. [Online]. Available: <https://prism.sustainability-directory.com/term/maritime-ai-ethics/>