

以查詢為中心的軌跡預測

Zikang Zhou^{1,2} Jianping Wang^{1,2} Yung-Hui Li³ Yu-Kai Huang⁴

¹香港城市大學²香港城市大學深圳研究院

³鴻海研究院⁴卡內基美隆大學

zikanzhou2-c@my.cityu.edu.hk

抽象的

預測周圍智能體的未來軌跡

對於自動駕駛汽車的安全運行至關重要。這
論文提出了 QCNNet，一個突破軌跡預測界限的建模框架。首先，我們確定現有的以主體為中心的建模方案

每當觀察視窗向前滑動時，這些方法就需要對輸入進行重新歸一化和重新編碼，從而導致在線預測期間出現冗餘計算。

為了克服這個限制並實現更快的推理，
我們引入了以查詢為中心的場景編碼範例，
透過學習可以重複使用過去的計算
獨立於全球時空座標系的表示。在所有人之間共享不變的場景特徵

目標代理進一步允許多代理的並行性
軌跡解碼。其次，即使給予豐富的編碼
現有的解碼策略難以捕捉的場景
代理人未來行為固有的多模態，特別是當預測範圍很長時。為了解決這個問題
挑戰，我們首先使用無錨查詢來生成
以循環方式提出軌跡建議，這使得
該模型在解碼不同地平線的路點時利用不同的場景情境。細化模組

然後以軌跡建議為錨點和槓桿
基於錨點的查詢進一步細化軌跡。經過
為細化模組提供自適應和高品質的錨點，我們基於查詢的解碼器可以
更好地處理
軌跡預測輸出的多模態。我們的
方法在 Argoverse 1 和 Argoverse 2 運動中排名第一
預測基準，在所有方面都優於所有方法
主要指標大幅提升。同時，我們的模型可以
實現串流場景編碼和平行多智能體
得益於以查詢為中心的設計理念的解碼。

一、簡介

為自動駕駛汽車做出安全決策需要
準確預測周圍智能體的未來軌跡。近年來，基於學習的方法

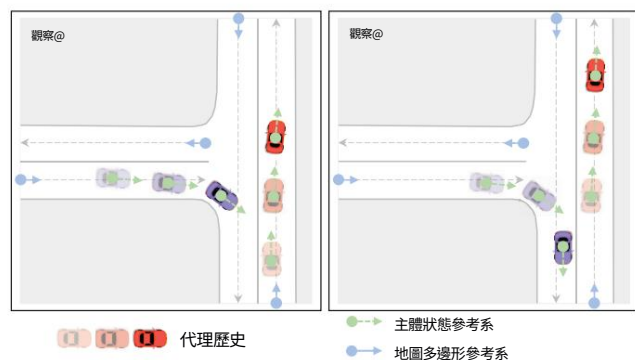


圖 1. 以查詢為中心的參考架構的圖示，
我們為每個時空建立一個局部座標系
元素，包括地圖多邊形和所有時間步驟的代理狀態。
在基於注意力的編碼器中，所有場景元素的查詢都是在其本地參考幀中匯出和更新的。

廣泛用於軌跡預測[14,31,37,38,46,56]。

儘管為增強模型的性能付出了巨大的努力
預測能力，離充分發揮還有很長的路要走
解決軌跡預測問題。為什麼是這樣
任務如此具有挑戰性，現有方法有何不足？我們試圖從以下方面回答這些問題
以下兩個觀點：

(i)雖然蓬勃發展的預測模型
在軌跡預測方面取得了令人印象深刻的表現
基準測試[7,13,49]，現今最先進的架構
專門用於此任務的[37,38,46,56]無法處理
有效地實現異質交通場景。在一個自治的
驅動系統，資料幀到達預測模組
依序作為稀疏場景上下文流，包括
高清向量圖及週邊特工
運動狀態。模型必須學習這些場景元素的表達表示才能實現準確的
預測。
隨著建模技術的不斷發展
對於稀疏上下文編碼[14,31,50]，研究界已經見證了朝著更強大的
方向快速進步
軌跡預測器。值得注意的是，基於分解的注意力
Transformers [37,38,56]最近將預測精度提高到了前所未有的水平。然而，他們要求
學習每個空間的基於注意力的表示

時間場景元素並遭受過高的影響
處理密集交通場景時的成本。因為每一個最小的延遲都可能導致自動駕駛中的災難性事故
駕駛時，即時預測的未滿足需求有限
最先進方法的適用性。

(ii)軌跡輸出的巨大不確定性
隨著預測範圍的延長，預測的爆發性成長一直困擾著研究界。例如，車輛在十字路口可能會轉向

或直行取決於駕駛員的長期目標。到
為了避免錯過任何潛在的行為，模型必須學會
捕獲潛在的多峰分佈而不是
只是預測最常見的模式。這次學習
任務具有挑戰性，因為只有一種可能性已登錄
每個訓練樣本。為了緩解學習困難，
的作品利用手工製作的錨作為多模態預測的指導[6,12,39,53,55]。他們的有效性，
然而，這取決於錨的品質。通常，
當很少有錨能夠精確覆蓋地面事實時，這些方法就無法很好地發揮作用。這個問題更加嚴重
在長期預測中，錨點的搜尋空間
更大。其他一些作品[10, 31, 38, 46, 56]透過直接預測多個軌跡來規避這個問題，儘管存在模式崩潰和訓練不穩定的風險[33, 41]。由於缺乏空間先驗，這些方法

也無法產生準確的長期預測。

上述分析促使我們提出一種軌跡預測框架，稱為 QCNet，以克服先前解決方案的限制。首先，我們注意到可以實現更快的線上推理，同時也受益

來自分解注意力的力量，但以主體為中心
現有方法所使用的編碼方案[25,27,46,56]
充當障礙。每次新的資料幀到達時，觀察視窗向前滑動一步，

與其前身實質上重疊，這提供了
模型重用先前計算的機會
編碼。然而，以代理為中心的方法需要根據最新代理狀態的位置對輸入進行標準化，
每當需要重新編碼場景元素
觀察窗向前滑動。為了解決這個問題，我們引入了一個以查詢為中心的場景編碼範例（見圖1）。我們設計精神的關鍵在於

處理本地時空參考系中的所有場景元素並學習獨立的表示

全局座標。該策略使我們能夠緩存
並重複使用先前計算的編碼，傳播
跨所有觀察視窗的計算，從而減少推理延遲。不變的場景特徵可以

也可以在場景中的所有目標代理之間共享，以實現多代理解碼的並行性。其次，為了
为了更好地利用場景編碼進行多模態和長期預測，我們使用無錨查詢來檢索
場景上下文循環並讓他們解碼一個短片段

每次重複時都會調整未來的航路點。這種反覆發作的
此機制允許查詢在預測不同地平線的路徑點時專注於不同的場景上下文，從而減輕查詢的建模負擔。高品質

循環解碼器預測的軌跡在後續細化模組中充當動態錨點，其中

我們使用基於錨點的查詢來根據場景上下文細化軌跡建議。因此，我們基於查詢的
解碼管道結合了無錨的靈活性
方法融入基於錨的解決方案，充分利用兩者的優點
促進多模式和長期預測的世界。

我們提出的以查詢為中心的編碼範例是
首先，可以利用軌跡預測的順序性來實現快速在線推理。此外，我們基於查詢的解碼器在多模式和長期預測方面表現出卓越的效能。實驗表明

我們的方法取得了最先進的結果，在
兩個大規模運動預測基準[7, 49]。

2.相關工作

場景上下文融合編碼豐富的資訊用於軌跡預測。早期工作將世界狀態光柵化為多通道影像，並採用經典的捲積神經網絡

學習網絡[5,6,10,21]。由於有損渲染、有限的感受野以及過高的成本
基於柵格的方法，研究界已經轉向
到基於向量的編碼方案[14,31,50]。隨著
使用排列不變集合運算符，例如池化[3, 12, 14, 20, 46]，圖卷積[11, 31, 36, 53]，
和注意力機制[24,26,30,32,34,52]，基於向量的方法可以有效地聚合交通場景中的稀疏資訊。幾個強大的軌跡預測

模型最近採用了具有分解注意力的Transformer [47]作為編碼器[18,37,38,56]。雖然
這些模型透過以代理人為中心的學習來提高效率
分層表示[56]或對整體進行編碼
共享座標系中的場景[38]，它們的可擴展性是
仍然受到因式分解的計算複雜性的限制
注意力。相較之下，我們的編碼器繼承了解析注意力的表示能力，同時實現了更多
透過使用以查詢為中心的編碼範例來實現高效的場景上下文融合，該範例超越了以代理為中心的建模並支援串流軌跡預測。

考慮到世界狀態，多模態未來分佈是一種廣泛採用的軌跡預測輸出形式
是部分可觀察的，代理人的意圖是高度
不確定。雖然生成模型自然適合多模態預測[20,28,40,45]，但從潛在變數中取樣會引入測試時隨機性，這對於自動駕駛等安全關鍵應用來說是不受歡迎的

駕駛。另一項研究透過以下方式解決多模態問題：
從編碼中解碼一組離散的軌跡
場景脈絡[6, 10, 31, 55]。由於只有一種模式是ob-

服務訓練數據，預測多種不同的未來
具有挑戰性。基於錨的方法透過以下方式實現這一點
錨點的引導，促進多模態預測
透過利用預先定義的操作[12]、候選軌跡[6, 39] 或地圖自適應目標
[53, 55]。然而，
這些錨點的品質顯著影響預測性能。相比之下，無錨方法輸出

自由地進行多種假設，冒著模式崩潰的風險
訓練不穩定[10, 31, 38, 46]。我們的解碼管道
利用基於錨點和無錨點的解決方案，無錨點模組以數據驅動的方式產生自適應錨點，以及基於錨點的模組

根據場景上下文細化這些錨點。

3. 方法

3.1.輸入和輸出公式

考慮 A 智能體圍繞自動駕駛車輛的場景。在線運行過程中，感知

模組為預測提供代理狀態流
模組以固定的間隔進行，其中每個代理狀態與其時空位置和幾何屬性
相關聯。例如，第 i 個智能體在時間步 t 的狀態

包括空間位置 $\mathbf{p}_t^i = (\mathbf{p}_t^i, \theta_t^i)$ ，角度
位置 θ_t^i （即偏航角）、時間位置 \mathbf{t} （即
時間步長）和速度 \mathbf{v}_t^i 。我們還添加了運動
向量點 \mathbf{p}_{t-1}^i 與某些類似的幾何屬性
基線[31, 56]。此外，預測模組可以存取高清地圖上的 M 個多邊形（例
如，車道）
和人行橫道），其中每個地圖多邊形都用
採樣點和語義屬性（例如，使用者類型
車道）。給定地圖資訊和代理狀態
在 T 個時間步長的觀察視窗內，預測模組的任務是預測 K 個未來軌
跡
對於 T 時間步長範圍內的每個目標代理，以及
為每個預測分配機率分數。

3.2.以查詢為中心的場景上下文編碼

軌跡預測的第一步是對
場景輸入。最近的研究發現分解注意力
對於場景編碼非常有效[37, 38, 56]。這些
方法讓查詢元素專注於鍵/值元素
一次沿著一個軸，這會導致暫時的注意力，
主體圖注意力與社會注意力（即主體-主體
注意），複雜度為 $O(AT_2)$ 、 $O(ATM)$ 和
分別為 $O(A_2T)$ 。與典型的編碼策略不同
首先應用時間網路來壓縮時間維度，然後僅在當前時間步執行智能體-
地圖和智能體-智能體融合，分解注意力在觀察內的每個過去時間步進
行融合

窗戶。因此，分解注意力可以捕獲更多
訊息，例如代理人之間的關係如何
地圖元素在觀察範圍內演變。如何-

曾經，它的可擴展性受到立方複雜度的限制
每次融合操作。在極端情況下涉及
數百個代理商和地圖元素，這樣的模型可能會失敗
及時發出預測。我們問：是否有可能減少線上預測期間的推理延遲，
同時
享受分解注意力的代表性力量？

在深入研究我們的解決方案之前，請記住軌跡預測是一項流處理
任務：當新數據
幀到達，我們將其放入隊列並丟棄最舊的
一。因此，最近的觀測窗口有 $T-1$ 時間
步驟與其前身重疊。這一事實激勵
我們提出另一個問題：我們可以重複使用重疊的部分嗎？
觀察視窗向前滑動後先前計算的時間步長編碼？不幸的是，這個想法

由於標準化要求，這是不可行的
軌跡預測：現有方法採用以代理為中心的編碼範例來實現空間旋轉平
移不變性[25, 27, 46, 56]，其中每個代理程式都編碼為

由目前時間決定的局部座標系
步的位置和偏航角。每次觀察
視窗向前滑動，「目前時間步長」也發生移動
因此，所有場景元素的幾何屬性都需要根據其位置重新歸一化

最新的代理人表示。由於輸入的變化，我們
被迫重新編碼所有時間步的元素，即使
觀察視窗大部分重疊。

基於上述分析，我們發現不斷演變的時空座標系阻礙了先前計算
的編碼的重複使用。為了解決這個問題，我們

引入以查詢為中心的學習編碼範例
獨立於場景元素的全域座標的表示。具體來說，我們為每個場景元素
建立一個局部時空座標系，查詢向量

派生自，處理查詢元素的特徵
局部參考系。然後，我們將相對時空位置注入到鍵和值元素中

執行基於注意力的場景上下文融合。我們在下面的段落中詳細闡述編
碼過程。

當地時空座標系。圖1顯示了
場景元素的局部座標系範例。為了
第 i 個智能體在時間步 t 的狀態，局部座標
框架由參考時空位置 \mathbf{pos}_i^t 和 θ_t^i 確定
化（點 \mathbf{t} ）和參考方向 θ_t^i ，其中 \mathbf{p}_t^i 和 θ_t^i
分別是主體狀態的空間位置和角度位置。對於車道和人行橫道，我們
選擇位置
和中心線入口點的方向作為
參考。這樣我們就建立了局部座標系
規範地考慮所有場景元素，結果
在任何觀察視窗內，每個地圖多邊形都有一個專用的本地
幀，每個智能體都有 T 參考幀。

場景元素嵌入。對於每個時空
場景元素，例如代理狀態或車道，我們計算所有幾何屬性的極座標（例
如，

代理狀態的速度和運動向量，車道上所有採樣點的位置)相對於空間

元素的本地框架引用的點和方向。
然後，我們將每個極座標轉換為傅立葉特徵[22,35,44]以方便學習高頻訊號。
對於每個代理狀態和每個採樣點

在地圖上，傅立葉特徵與
語意屬性（例如，代理的類別）並傳遞
透過多層感知器（MLP）來獲得嵌入。進一步產生多邊形級表示

對於車道和人行橫道，我們對每個地圖內採樣點的嵌入應用基於注意力的池化
多邊形。這些操作導致代理嵌入
形狀 [A, T, D] 和形狀 [M, D] 的地圖嵌入，
其中 D 表示隱藏特徵維度。受益於局部參考系中的建模，嵌入

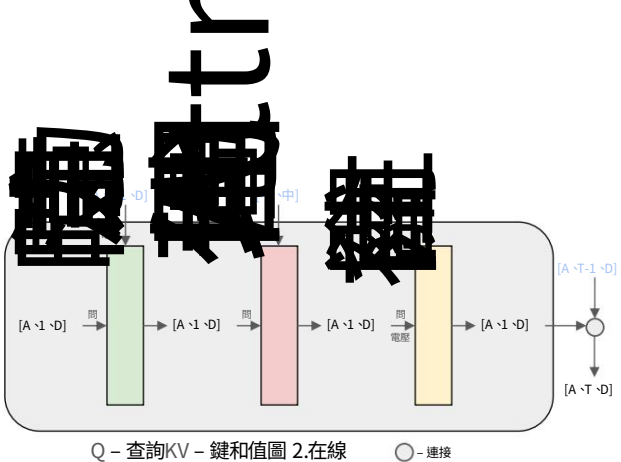
每個代理狀態/地圖多邊形只有一個實例並且
可以在後續的觀察視窗中重複使用。在
相比之下，以代理為中心的方法必須複製所有輸入
多次，相對於 A 之一對每個副本進行編碼
代理的當前位置和航向，並在觀察視窗向前滑動時重新編碼所有輸入，從而導致
在線推理期間產生更多開銷。

相對時空位置嵌入。我們
為場景元素對準備相對位置嵌入，該嵌入將合併到基於注意力的算子中，以幫助
模型了解兩個元素的局部座標系之間的差異。為

具有絕對時空位置的元素 (pt_{我, t})
另一個帶有 (ps_{那個j, t}, s)，我們使用 4D 描述子來
總結它們的相對位置，其組成部分是
相對距離ps_{我, t}，點我，相對方向
jatan2(ps_{我, t}, y-pt_{我, t}, x-pt_{我, t}) - θ_{我, t}，相對方向
那個j, t, y, x和時間間隔 s - t。因為我們可以輕鬆地重新
構造一個元素相對於另一個元素的絕對位置
在描述符的幫助下，我們保留了場景元素對的所有時空位置資訊。

然後，我們將 4D 描述子轉換為傅立葉特徵
並將它們傳遞給 MLP 以產生相對位置嵌入rs→t如果兩個場景元素中的任何
一個j→i。
是靜態的（例如，靜態地圖多邊形），我們可以省略上標並將嵌入表示為
rj→i。
地圖編碼的自注意力。我們利用自注意力來建模地圖元素之間的關係，

之後更新的地圖編碼將豐富代理
特徵並協助軌跡解碼。對於第 i 個地圖
多邊形，我們從其嵌入mi導出查詢向量
讓它注意鄰近的車道和行人穿越道
{mj}j∈Ni，其中Ni表示多邊形的鄰居集合。為了將空間意識納入地圖編碼中，
我們
從串聯產生第 j 個鍵/值向量
mj和相對位置嵌入，即 [mj, rj→i]。
由於(mi, mj, rj→i)的每個三元組輸入到注意力



模式下的編碼器概述。在重複使用之前觀察視窗（藍色）中計算的編碼後，
分解注意力的複雜性從

O(AT²)+O(ATM)+O(A2T)到 O(AT)+O(AM)+O(A2)。

層獨立於全域時空座標系，在全域參考系變換下輸出圖編碼{m_i}M i=1。因此，
也是不變的

它們可以在所有代理和所有時間步之間共享，
甚至可以離線預先計算，從而避免以代理為中心的建模所遭受的冗餘計算。

代理編碼的分解注意力。為了幫助
代理嵌入捕獲更多信息，我們還考慮跨代理時間步驟的分解注意力，其中

代理商以及代理商和地圖之間，採取第 i 個代理
以時間步t為例。給定匯出的查詢向量
來自代理狀態的嵌入，我們使用時間
注意力透過計算鍵和值向量基於
[[作義; r_{s→t}]] s=t-τ，這是第 i 個智能體的嵌入
從時間步 t-τ (0<τ<T) 到時間步 t-1 以及對應的相對位置嵌入。同樣，

主體地圖和社會關注的鍵和值向量
源自 [m_j, r_{j→i}] j∈Ni 和在 [j; r_{t→tj}→i] j∈Ni，
分別，其中鄰居集合Ni由 a 決定
距離閾值為 50 公尺。由於更新了
最初具有不變鍵和值的不變查詢，
這些層的輸出也是不變的。我們依序堆疊時間、主體圖和社會注意力

作為一個融合塊並重複此類塊Lenc次。
由於以查詢為中心的建模，所有代理和
地圖編碼是唯一且固定的，無論來自哪個
我們查看它們的時空座標系（即空間維度的旋轉平移不變性和時間維度的
平移不變性），使模型成為可能

重複使用過去的計算並進行串流操作。在線上預測過程中，我們可以快取計
算出的編碼
在先前的觀察視窗中並逐步更新
場景表示。如圖2所示，我們的模型
只對 A 傳入的智能體執行分解注意力
狀態新資料幀到達時，導致時間
具有 O(AT) 複雜度的注意力，具有 O(AT) 複雜度的注意力，具有
O(AM) 複雜度，以及 O(A2)複雜度的社會關注。所有這些操作都更便宜

比他們的非串流媒體對手。最後，我們更新一下
使用新計算的編碼快取張量。

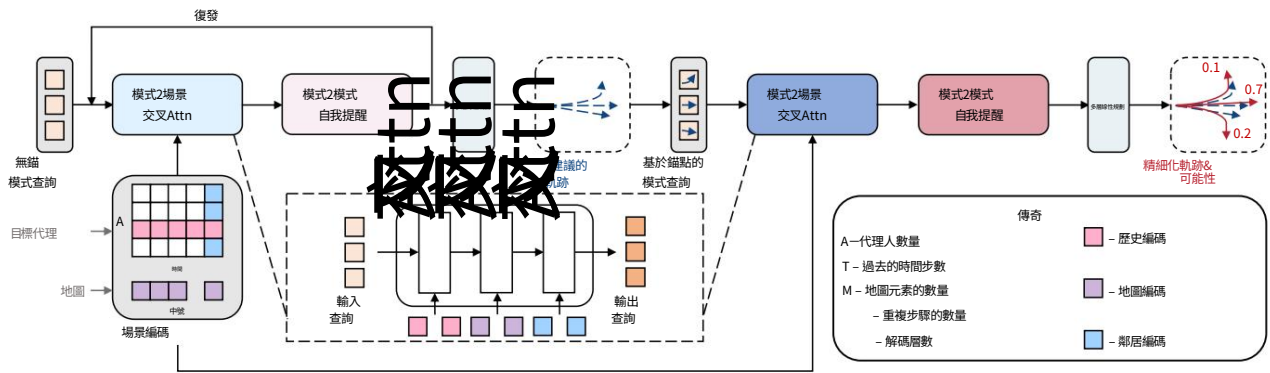


圖 3. 解碼管道概述。無錨模組根據編碼循環產生軌跡建議
場景上下文。這些建議充當細化模組中的錨點，其中基於錨點的解碼器細化錨點軌跡
並為每個假設分配一個機率分數。

3.3.基於查詢的軌跡解碼

軌跡預測的第二步是利用編碼器輸出的場景編碼，用於解碼每個目標代理程式的 K 個未來軌跡，這非常重要。因為編碼器僅傳回一組特徵嵌入。受目標偵測進展的啟發，最近的一些工作 [18,32,37,46] 採用了類似 DETR 的解碼器 [4]。

處理這樣一個一對多的問題，其中多個可學習的查詢交叉參與場景編碼和解碼軌跡。然而，與其他無錨方法一樣，這些模型也存在訓練不穩定和模式崩潰的問題。此外，它們在長期預測中表現不佳，因為預測任務要多得多。

由於遙遠的爆炸性不確定性而具有挑戰性未來。我們基於查詢的解碼器透過利用循環的無錨提議模組來產生自適應軌跡錨點，然後透過

基於錨點的模組，進一步細化初始提案。我們的解碼管道的概述如圖所示。

圖3. 下面我們將說明各個組件解碼器的詳細介紹。

Mode2Scene 和 Mode2Mode Attention。提案和細化模組都使用類似 DETR 的架構。與 DETR [4] 中物件查詢的概念類似，

每個查詢負責解碼 K 種軌跡模式之一。在 Mode2Scene 注意力中，我們使用交叉注意力層來更新具有多個的模式查詢。

上下文，包括目標代理的歷史編碼，地圖編碼和鄰近代理人的編碼。繼 Mode2Scene 注意力之後， K 模式查詢透過 Mode2Mode self-attention 相互“交談”。

提高多種模式的多樣性。模式查詢的參考系。為了並行預測多個智能體的軌跡，我們共享相同的場景中所有目標代理程式之間的場景編碼集。由於這些編碼源自其本地時空座標系，我們需要將它們投影到每個目標代理的當前視點中以達到與

以主體為中心的建模。為此，我們根據對應目標代理程式的當前位置和偏航角為每個模式查詢幻覺一個座標系。什麼時候

透過 Mode2Scene 注意力更新查詢嵌入，場景元素相對於查詢的位置被合併到鍵和值中，這類類似於

我們已經為編碼器做了。

無錨軌跡提案。我們使用可學習的、無錨查詢來提出初始軌跡。這些提案稍後將充當細化模組中的錨點。與嘗試基於錨的方法相比

用密集採樣的手工製作來掩蓋事實真相錨 [6, 19]，我們的提案模組以數據驅動的方式產生 K 個自適應錨。由於交叉注意層，模式查詢可以檢索場景

上下文並快速縮小錨點的搜尋空間。自注意力層還允許查詢在產生軌跡建議時相互協作。

在延長的預測範圍內，代理商可以旅行距離較遠，周圍環境可能有所不同迅速地。因此，很難概括所有信息。將長序列解碼為單一查詢所需的嵌入。為了減輕查詢的上下文提取負擔並提高錨點的質量，我們概括了類似 DETR 的解碼器成為一種流行的方式。使用 Trec 循環步驟，上下文感知模式僅查詢解碼 T/Trec 在每個循環步驟結束時透過 MLP 確定未來航路點。在隨後的重複中，這些查詢再次成為輸入並提取與接下來幾個路點預測相關的場景上下文。為了效率，Trec 遠小於預測範圍 T 。我們也發現使用更多的重複步驟是不必要的。基於錨點的軌跡細化。無錨解碼可以是一把雙面刃：儘管它具有靈活性，不穩定的訓練過程有時可能會導致模式崩潰。另一方面，隨機初始化模式查詢必須適應所有場景中的所有目標代理，並且缺乏特定場景的偏差，這可能會導致非

合規的預測。例如違反規則的軌跡
運動規律或違反交通規則
高清图。因此，我們有動力僱用
基於錨點的模組進一步細化提案。將提案模組的輸出作為錨點，我們讓

細化模組預測所提出軌跡的偏移並估計每個假設的可能性。這

模組也採用了類似DETR的架構，但其模式
查詢是從建議的軌跡錨點導出的，而不是隨機初始化的。具體來說，一個小型 GRU [8]
用於嵌入每個軌跡錨點，我們取其最終的
隱藏狀態作為模式查詢。這些基於錨點的查詢
為模型提供明確的空間先驗，使注意力層更輕鬆地定位感興趣的上下文。

3.4. 培訓目標

遵循 HiVT [56]，我們參數化第 i 個代理的
未來軌跡作為拉普拉斯分佈的混合：

$$f(\{p_t\}_{t=1}^K) = \prod_{k=1}^K \pi_{i,k} \text{ 拉普拉斯點 } \mu_{i,k}, \sigma_{i,k} \quad (1)$$

其中 $\{\pi_{i,k}\}_{k=1}^K$ 是混合係數，第 k 個
時間步 t 處混合物成分的拉普拉斯密度由位置 $\mu_{i,k}$ 參數化
和規模 $\sigma_{i,k}$ 。然後我們
使用分類損失 L_{cls} 來優化細化模組預測的混合係數。這種損失最小化
了方程式的負對數似然。(1)，我們停止

位置和比例的梯度以優化混合
僅係數。另一方面，我們採用贏者全拿的策略 [29] 來優化地點和規模

提案和細化模組的輸出，對最佳預測的提案進行反向傳播，

僅其精煉。為了穩定，細化模組停止了所提出的軌跡錨的梯度。

最終的損失函數結合了軌跡提議損失
 $L_{propose}$ 、軌跡細化損失 L_{refine} 和端到端訓練的分類損失 L_{cls} ：

$$L = L_{propose} + L_{refine} + \lambda L_{cls}, \quad (2)$$

我們使用 λ 來平衡回歸和分類。

4. 實驗

4.1. 實驗設定

數據集。我們使用 Argoverse 1 [7] 和 Argoverse 2 [49]，
兩個大規模運動預測資料集，以測試我們方法的有效性。Argoverse
1 資料集收集
323、557 個來自邁阿密和匹茲堡的資料序列，
而 Argoverse 2 資料集包含 250,000 個場景
橫跨六座城市。兩個資料集的取樣率為
10赫茲。對於 Argoverse 1 資料集，模型需要在給定 2 秒的情況下預測
智能體 3 秒的未來軌跡

對歷史的觀察。相較之下，Argoverse 2 資料集的特點是資料多樣性提
高、資料量更高
質量、5 秒的更大觀察窗口，以及
6 秒的更長預測範圍。使用這兩個
資料集，我們打算檢查模型對各種資料分佈和預測範圍的預測能力。

指標。遵循標準評估協議，我們
採用包括最小平均位移在內的指標
誤差 (minADEK)，最小最終位移誤差
(minFDEK)、Brier-最小最終位移誤差 (b-minFDEK) 和缺失率 (MRK)
進行評估。metric minADEK 計算之間的距離（以公尺為單位）
 $\frac{1}{2}$

地面實況軌跡和 K 個預測軌跡中的最佳軌跡作為所有未來時間步的平
均值。上
另一方面，指標 minFDEK 僅關注最後時間步的預測誤差，以強調長期

表現。為了進一步衡量不確定性估計的性能，度量 b-minFDEK 添加了
 $(1 - \pi)^2$
到最後一步誤差，其中 π 表示最佳預測
模型分配的軌跡機率得分。此外，度量 MRK 用於計算案例的比例

其中 minFDEK 超過 2 公尺。作為一種常見的做法，
 K 選為 1 和 6。如果模型輸出多於 K
軌跡，僅具有前 K 個機率的預測
評估過程中會考慮分數。

4.2. 與現有技術的比較

我們將我們的方法與最強的基線進行比較
Argoverse 1 和 Argoverse 2 運動預測
基準 [7, 49]。我們先進行實驗
Argoverse 2 資料集 [49]，支援有效的解決方案
由於其預測範圍長達 6 秒，因此在長期預測方面表現良好。結果如表所
示。1。
即使沒有集成，QCNet 的性能已經優於
Argoverse 2 測試集上的所有先前方法
minADE6、minFDE6、minADE1 和 minFDE1。後
使用與其他條目類似的整合技術，QC-Net 在所有指標上都大幅超越了
所有方法。
我們也在 Argoverse 1 資料集 [7] 上評估我們的模型
更好地理解我們方法的普遍性。儘管 Argoverse 1 基準測試的效能有
所下降
飽和多年 [49]，表。圖 2 顯示 QCNet 在大多數指標上都顯著提升了最
先進水準。截至
當我們提交論文時，QCNet 在 Argoverse 1 和 Argoverse 2
的排行榜上排名第一，表現優於所有其他
關於這兩個基準的已發表和未發表的作品。
更多結果請參考補充資料
Argoverse 2 [49] 和 Waymo 開放運動資料集 [13]。

4.3. 消融研究

場景情境融合的效果。我們研究的影響
Tab 中的場景上下文融合。3. 我們要回答的第一個問題是分解注意力
是否值得。如果沒有事實

方法	b-minFDE6 ↓	minADE6 ↓	minFDE6 ↓	MR6 ↓	minADE1 ↓	minFDE1 ↓	MR1 ↓
湯瑪斯[17]	2.16	0.88	1.51	0.20	1.95	4.71	0.64
戈熱拉[9]	2.01	0.76	1.48	0.22	1.82	4.62	0.66
港鐵[42]	1.98	0.73	1.44	0.15	1.74	4.39	0.58
GANet [48]	1.96	0.72	1.34	0.17	1.77	4.48	0.59
QML* [43]	1.95	0.69	1.39	0.19	1.84	4.98	0.62
BANet* [54]	1.92	0.71	1.36	0.19	1.79	4.61	0.60
QCNet (無整體)	1.91	0.65	1.29	0.16	1.69	4.30	0.59
QCNet (附整合)	1.78	0.62	1.19	0.14	1.56	3.96	0.55

表 1.以 b-minFDE6 排名的Argoverse 2運動預測排行榜[1]的定量結果。已知的基線使用過的整合均標有符號 “*” 對於每個指標，最佳結果以粗體顯示，次佳結果以下劃線顯示。

方法	b-minFDE6 ↓	minADE6 ↓	minFDE6 ↓	MR6 ↓
LaneGCN [31]	2.06	0.87	1.36	0.16
mmTransformer [32]	2.03	0.84	1.34	0.15
密集TNT [19]	1.98	0.88	1.28	0.13
食品補充劑[50]	1.93	0.82	1.24	0.13
場景轉換器[38]	1.89	0.80	1.23	0.13
回家+回家[15, 16]	1.86	0.89	1.29	0.08
高電壓[56]	1.84	0.77	1.17	0.13
多路徑++ [46]	1.79	0.79	1.21	0.13
GANet [48]	1.79	0.81	1.16	0.12
支付[11]	1.76	0.80	1.21	0.11
DCMS [51]	1.76	0.77	1.14	0.11
引路者[37]	1.74	0.77	1.16	0.12
我們的	1.69	0.73	1.07	0.11

表 2. Argoverse 1運動預測排行榜的定量結果[2]。排行榜按b-minFDE6排序。

模型	在線推理 (ms)	minADE6 ↓	minFDE6 ↓	MR6 ↓	不重用	不重用
QCNet (Lenc = 0) 8 ± 1		1 ± 0	0.76	1.33	0.18	
QCNet (Lenc = 1) 64 ± 1		10 ± 1	0.74	1.30	0.17	
QCNet (Lenc = 2) 82 ± 1		13 ± 1	0.73	1.27	0.16	

表 3. 評估模型的效能和推理延遲
Argoverse 2 驗證集。我們使用A40 GPU來測量
編碼器在涉及190 個代理程式和169 個地圖多邊形的最密集交通
場景中的線上推理延遲。

採用基於注意力的融合區塊，編碼器不涉及沿著場景不同軸的資訊互動。即使如此，解碼器仍然可以訪問

得益於預測所需的所有信息
Mode2Scene 圖層。如Tab第一行所示。3、
我們的模型可以提供可靠的預測性能，而無需
融合場景上下文。但是在給我們的解碼器之後
透過融合編碼，該模型比其非融合變體獲得了相當大的改進。此外，增加
融合塊的數量在所有指標上都產生了更好的結果，證明了分解注意力的
有效性。
然而，由此產生的推理延遲不適合
自動駕駛等即時應用
這種模型選擇對於廣泛採用的吸引力較小。幸運的是，我們以查詢為中心的範式允許
重用以前觀察視窗的計算
在線上預測期間，這超出了以代理為中心的範圍
方法如 HiVT [56]。如表所示。3、緩存

並徹底重複使用先前計算的編碼
降低最密集流量下的線上推理延遲
場景。這樣的「免費午餐」也為設計更先進的解碼器以實現更強的性能
提供了空間。
解碼器的組件研究。如圖所示
標籤。如圖4所示，解碼器中的所有層都在一定程度上對效能
有貢獻。首先，雖然因式分解
編碼器中的注意力已經為代理特徵帶來了上下文感知，我們發現使用目
標
僅代理的歷史編碼不足以準確
軌跡建議。我們假設注入地圖
將社交資訊輸入到無錨解碼器中可以明確地為查詢提供未來的上下
文，從而使
它們縮小了初始軌跡的搜尋空間。在
另一方面，地圖和社交資訊也能提供幫助
基於錨點的查詢可以識別那些不切實際的預測，例如違反交通規則或
與道路上的靜態物體相撞的軌跡。結果，角色

我們的細化模組不僅僅是簡單的軌跡
平滑[23, 51]。表4也表明，刪除
無論是提案還是細化模組中的Mode2Mode自注意力都會損害長期準
確性和
多重假設的多元性。
表5展示了循環機制和細化模組對不同資料集的影響。在

Argoverse 1，大多數智能體僅表現出微不足道的行為，
且場景上下文通常在 3 秒預測範圍內沒有顯著變化。因此，我們的設計
選擇只能帶來邊際改進

在此資料集上進行評估時，然而，在更具挑戰性的 Argoverse 2 資料
集上，預測範圍為
6秒，將循環步驟數從1增加
(即，不重複)到 3 會帶來更好的長期性能，而細化模組在準確性和多
模態方面提供了顯著的改進。我們

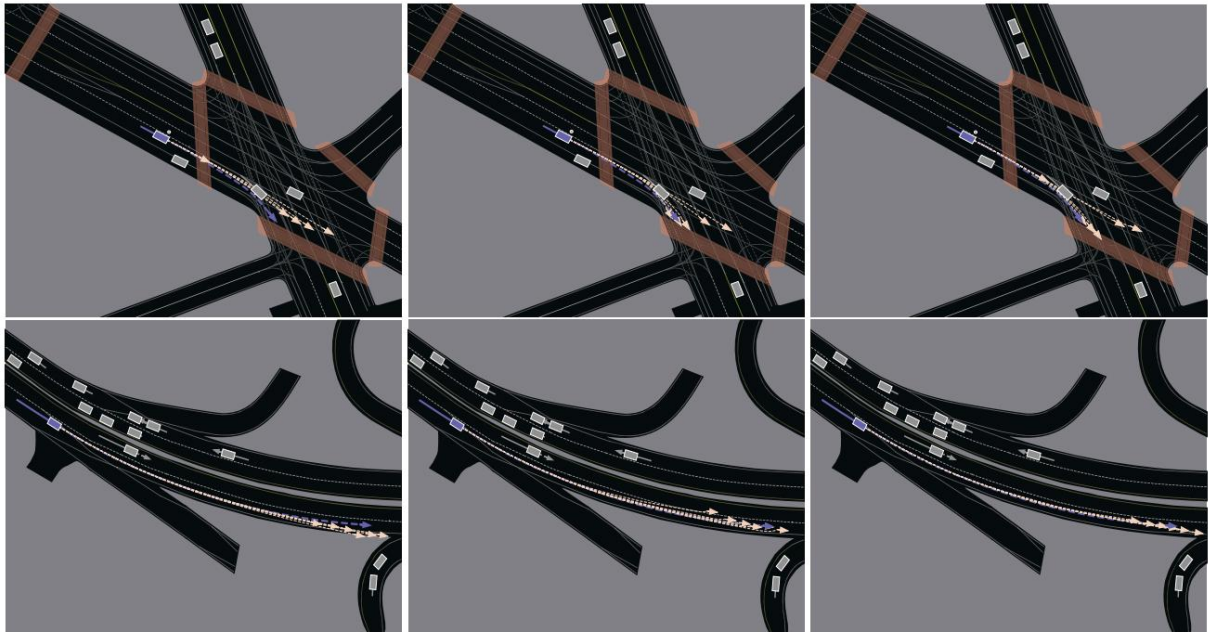
也請注意，使用更多的循環步驟是多餘的：當數量從 3 增加到 6 時，模
型
Argoverse 2 的效能無法進一步提升。

4.4.定性結果

我們展示了 Argoverse 2 的一些定性結果
驗證集。比較圖4a和圖4b，我們可以看到

提案模組	精煉模組	b-minFDE6 ↓ minADE6 ↓ minFDE6 ↓ MR6 ↓			
時間圖社交模式 時間圖社交模式					
		2.22	0.82	1.58	0.22
		2.04	0.78	1.43	0.19
		2.18	0.81	1.55	0.22
		2.06	0.79	1.48	0.21
		2.02	0.77	1.40	0.19
		1.99	0.74	1.33	0.17
		1.94	0.74	1.30	0.17
		1.97	0.74	1.31	0.17
		1.91	0.73	1.29	0.17
		1.90	0.73	1.27	0.16

表 4. 解碼器組件的消融研究。實驗結果基於Argoverse 2驗證集。



(a) 不帶遮障且不細化 (b) 帶遮障和不細化 (c) 帶遮障和帶細化

圖 4.Argoverse 2 驗證集的定性結果。顯示目標代理人的邊界框和地面實況軌跡。紫色顯示，模型預測顯示粉紅色。

數據集	#循環步驟細化	b-minFDE6 ↓	minFDE6 ↓	MR6 ↓	
Argoverse 1 2 (1.5 秒/步) (3 秒預測)	1 (3秒/步)	×	1.58	0.92	0.09
	3 (1秒/步) 3 (1秒/步)	×	1.57	0.90	0.08
		×	1.56	0.90	0.08
			1.55	0.89	0.08
Argoverse 2 2 (3秒/步) (6秒預測)	1 (6秒/步)	×	2.10	1.47	0.20
	3 (2秒/步) 6 (1秒/步)	×	2.04	1.42	0.19
		×	2.02	1.40	0.19
			1.90	1.27	0.16
			1.90	1.27	0.16

表 5. 軌跡建議和細化模組的效果。在具有不同難度等級和預測範圍的資料集上。

提案模組的循環機制可以減少長期預測誤差。圖4c進一步證明了細化模組的有效性，它提高了多個假設的多樣性，以及預測軌跡的平滑度。

5. 結論

本文介紹了 QCNet，一種神經架構，克服了軌跡預測中的一些重要挑戰。在以查詢為中心的建模設計理念的支援下，QCNet 保持了以下表示能力：

分解注意力，同時享受更快的推理速度。它透過採用循環、無錨軌跡建議模組來實現多模態和長期預測。

基於錨點的細化模組。QCNet 在大規模軌跡預測資料集上表現出前所未有的效能，證明了其設計的有效性。

致謝

這項工作得到了香港研究資助局 GRF 11200220、《科學與技術》的部分支持。深圳市科技創新委員會基金會。授予號：JCYJ20200109143223052。

參考

[1] Argoverse 2 :運動預測競賽。 <https://eval.ai/web/challenges/challenge-page/1719/概述> 造訪時間 :2022 年 11 月 11 日。 7

[2] Argoverse運動預測競賽。 <https://eval.ai/web/challenges/challenge-page/454/概述> 造訪時間 :2022 年 11 月 11 日。 7

[3] 亞歷山大·阿拉希·克拉塔斯·戈爾·維涅什·拉馬納坦，
亞歷山大·羅比凱·李飛飛和西爾維奧·薩瓦雷斯。 Social lstm :擁擠空間中的人體軌跡預測。
IEEE 電腦視覺會議論文集
和模式識別 (CVPR) , 2016 年。 2

[4] 尼可拉斯·卡里昂·法蘭西斯科·馬薩·加布里埃爾·辛奈夫·尼可拉斯
烏蘇尼爾·亞歷山大·基里洛夫和謝爾蓋·扎戈魯伊科。使用
Transformer 進行端對端物件偵測。在訴訟程序中
歐洲電腦視覺會議 (ECCV) ,2020。 5

[5] Sergio Casas ·羅文傑·Raquel Urtasun ·意圖網 :學習從原始感測器資料預測
意圖。在
機器人學習會議 (CoRL) 會議記錄，
2018。 2

[6] Yuning Chai ·Benjamin Sapp ·Mayank Bansal 和 Dragomir
安格洛夫。多路徑 :用於行為預測的多個機率錨定軌跡假設。在訴訟程序中

機器人學習會議 (CoRL) ,2019。 2, 3, 5

[7] 張明芳 ·John Lambert ·Patsorn Sangkloy ·Jag-jeet Singh ·
Slawomir Bak ·Andrew Hartnett ·De Wang ·Peter
卡爾·西蒙·盧西·德瓦·拉馬南等。阿爾戈宇宙 :3d
使用豐富的地圖進行追蹤和預測。在訴訟程序中
IEEE/CVF 電腦視覺與模式會議
認可 (CVPR), 2019。 1, 2, 6

[8] Kyunghyun Cho, Bart van Merriënboer, C. aglar Gulc, ehre, i
Dzmitry Bahdanau ·Fethi Bougares ·Holger Schwenk 和
約書亞·本吉奧。使用 rnn 學習短語表示
用於統計機器翻譯的編碼器-解碼器。在親-
2014 年實證方法會議論文集
自然語言處理 (EMNLP) , 2014。 6

[9] Alexander Cui ·Sergio Casas ·Kelvin Wong ·Simon Suo 和
拉克爾·烏爾塔松。Gorela :相對於觀點不變
運動預測。arXiv 預印本 arXiv:2211.02545, 2022。 7

[10] Henggang Cui, Vladan Radosavljevic, Fang-Chieh Chou,
Tsung-Han Lin ·Thi Nguyen ·Tzu-Kuo Huang ·Jeff Schnei-der 和
Nemanja Djuric。使用深度卷積網路進行自動駕駛的多模態軌跡預測。IEEE
國際會議論文集

機器人與自動化 (ICRA) ,2019。 2, 3

[11] 方達·張宇·高清晰度感知圖注意力
運動預測中的地圖。在 IEEE 國際機器人與自動化會議 (ICRA) 會議記錄中，

2022年2月7日

[12] Nachiket Deo 和 Mohan M Trivedi。卷積社交
車輛軌跡預測的池化。訴訟中
IEEE 電腦視覺與模式會議
認可研討會 (CVPRW), 2018。 2, 3

[13] Scott Ettinger, Shuyang Cheng, Benjamin Caine, Chenxi
Liu, Hang Zhao, Sabeek Pradhan, Yuning Chai, Ben Sapp,
查爾斯·R·齊·殷周·等。大規模互動運動

自動駕駛預測 :waymo 開放運動資料集。IEEE/CVF 國際會議論文集

電腦視覺會議 (ICCV), 2021。 1, 6

[14] Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir
安格洛夫·李聰聰和科迪莉亞·施密德。向量網：
從向量化表示中編碼高清地圖和代理動態。IEEE/CVF 會議論文集

計算機視覺與模式識別 (CVPR) ,2020.1 ,
2

[15] 湯瑪斯·吉爾斯·斯特凡諾·薩巴蒂尼·德茲米特里·齊什庫·博格丹
史坦丘列斯庫和法比安·穆塔德。主頁 :用於未來運動估計的熱圖輸出。IEEE
會議錄
智慧交通系統國際會議 (ITSC) ,2021。 7

[16] Thomas Gilles ·Stefano Sabatini ·Dzmitry Tsishkou ·Bog-dan
Stanculescu 和 Fabien Moutarde。Gohome :面向圖形的熱圖輸出，用於
未來的運動估計。
IEEE 國際會議論文集
機器人與自動化 (ICRA) ,2022。 7

[17] 托馬斯·吉爾斯·斯特凡諾·薩巴蒂尼·德茲米特里·齊什庫·博格丹·
斯坦丘列斯庫和法比安·穆塔德。托馬斯 :軌跡
具有學習的多代理採樣的熱圖輸出。國際學習代表會議 (ICLR) 會議記錄，
2022 年。 7

[18] 羅傑·吉爾吉斯·佛洛里安·戈萊莫·費利佩·科德維拉·馬丁
韋斯·吉姆·奧爾登·迪索薩·薩米拉·亞伯拉罕·卡胡·菲利克斯
海德和克里斯多福·帕爾。潛在變數序列集
用於聯合多智能體運動預測的變壓器。國際學習代表會議 (ICLR) 會議記錄，
2022。 2, 5

[19] Junru Gu, Chen Sun, and Hang Zhao. Densetnt: End-to-end
從密集的目標集進行軌跡預測。在訴訟程序中
IEEE/CVF 電腦視覺國際會議
(ICCV), 2021。 5, 7

[20] 阿格里姆古普塔·賈斯汀約翰遜·李飛飛·西爾維奧薩瓦雷斯·
和亞歷山大·阿拉希·社會甘 :具有生成對抗網路的社會可接受的軌跡。IEEE
電腦視覺與模式會議論文集

認可 (CVPR) ,2018 年。 2

[21] 喬伊·洪·班傑明·薩普和詹姆斯·菲爾賓。規則
道路 :使用卷積模型預測駕駛行為
的語義交互。IEEE/CVF 會議記錄
電腦視覺與模式辨識會議
(CVPR), 2019。 2

[22] 安德魯傑格爾·菲利克斯吉梅諾·安迪布洛克·奧裡奧爾維尼亞爾斯·
安德魯·齊瑟曼和若昂·卡雷拉。感知者 :一般
具有迭代注意力的感知。國際機器學習會議 (ICML) 論文集 ,2021 年。

4

[23] Xiaosong Jia, Li Chen, Penghao Wu, Jia Zeng, Junchi Yan,
李紅陽，於喬。捕捉軌跡預測的時間動態 :一種從粗到細的方法。機器人學習會
議論文集

(CoRL) ,2022 年。 7

[24] Xiaosong Jia, Liting Sun, Masayoshi Tomizuka, and Wei
詹。Ide-net :互動式駕駛事件和從人類資料中提取模式。IEEE 機器人與自
動化 Let-
三 (RA-L) ,2021。 2

[25] Xiaosong Jia, Liting Sun, Hang Zhao, Masayoshi Tomizuka, 還有魏湛。透過結合自我中心和異中心觀點進行多智能體軌跡預測。在訴訟程序中
機器人學習會議 (CoRL) 2021. 2, 3

[26] Xiaosong Jia, Penghao Wu, Li Chen, Hongyang Li, Yu Liu, 還有嚴俊馳。Hdgt: 透過場景編碼進行多智能體軌跡預測的異構驅動圖轉換器。arXiv 預印本 arXiv:2205.09753, 2022. 2

[27] Miltiadis Kofinas、Naveen Nagaraja 和 Efstratios Gavves。用於交互動態系統的旋轉翻譯局部座標系。神經資訊處理系統進展 (NeurIPS) 2021. 2, 3

[28] Namhoon Lee、Wongun Choi、Paul Vernaza、Christopher B Choy、菲利普·HS 托爾和曼莫漢·錢德拉克。慾望：互動動態場景中的遙遠未來預測代理。IEEE 電腦會議論文集
視覺與模式識別 (CVPR) , 2017 年。2

[29] 史蒂芬李、森蒂爾普魯斯瓦卡姆希瓦普拉卡什、邁克爾科格斯韋爾、維雷什·蘭詹、大衛·克蘭德爾和德魯夫·巴特拉。用於訓練多樣化深度的隨機多項選擇學習合奏團。神經資訊處理進展系統 (NIPS) , 2016 年。6

[30] Lingyun Luke Li, Bin Yang, Ming Liang, Wenyuan Zeng, 任夢野、肖恩·西格爾和拉克爾·烏爾塔松。端到端交互變壓器的情境感知與預測。2020 年 IEEE/RSJ 國際智慧機器人與系統會議 (IROS) 論文集。2

[31] Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song 馮和拉克爾·烏爾塔松。學習運動預測的車道圖表示。在歐洲議會訴訟中
電腦視覺會議 (ECCV), 2020. 1, 2, 3, 7

[32] Yicheng Liu, Jinghui Zhang, Liangji Fang, Qinhong Jiang, 和周伯雷。堆疊式多模態運動預測變壓器。IEEE/CVF 會議論文集
計算機視覺和模式識別 (CVPR) 2021 年。2-5, 7, -

[33] Osama Makansi、Eddy Ilg、Ozgun Cicek 和 Thomas Brox。克服混合密度網絡的限制：多模式未來預測的取樣和擬合框架。
IEEE/CVF 電腦會議論文集
視覺與模式識別 (CVPR) , 2019 年。2

[34] Jean Mercat、Thomas Gilles、Nicole El Zoghby、Guil-laume Sandou、Dominique Beauvois 與 Guillermo Pita 吉爾。用於多模態聯合車輛運動預測的多頭注意力。IEEE 國際會議論文集
機器人與自動化會議 (ICRA) 2020. 2

[35] 本·米爾登霍爾、普拉圖·P·斯里尼瓦桑、馬修·坦西克、喬納森·T·巴倫 (Jonathan T Barron)、拉維·拉馬莫爾蒂 (Ravi Ramamoorthi) 和吳仁 (Ren Ng)。將場景表示為用於視圖合成的神經輻射場。歐洲電腦視覺會議 (ECCV) 會議記錄 2020 年。4

[36] 阿卜杜拉·穆罕默德·錢坤、穆罕默德·埃爾霍賽尼、克里斯蒂安·克洛岱爾。Social-stgcnn: 社會時空用於人類軌跡的圖卷神經網絡
預言。IEEE/CVF 會議論文集
計算機視覺和模式識別 (CVPR) , 2020。2

[37] Nigamaa Nayakanti、Rami Al-Rfou、Aurick Zhou、Kratarth 戈埃爾、哈立德·S·雷法特和班傑明·薩普。弓路者：透過簡單且有效率的注意力網絡進行運動預測。
arXiv 預印本 arXiv : 2207.05844 , 2022.1,2,3,5,7

[38] Jiquan Ngiam, Benjamin Caine, Vijay Vasudevan, Zheng-dong Zhang, Hao-Tien Lewis Chiang, Jeffrey Ling, Rebecca Roelofs、Alex Bewley、Chenxi Liu、Ashish Venugopal 等。場景轉換器：用於預測多智能體軌跡的統一架構。國際會議
學習表徵會議 (ICLR) 2022 年。1-2、3-7、-

[39] Tung Phan-Minh、Elena Corina Grigore、Freddy A Boulton、奧斯卡·貝伊博姆和埃里克·M·沃爾夫。覆蓋網：多式聯連使用軌跡集進行行為預測。在訴訟程序中
IEEE/CVF 電腦視覺與模式會議
認可 (CVPR), 2020. 2, 3

[40] 尼可拉斯·萊因哈特、羅文·麥卡利斯特、克里斯·基塔尼和謝爾蓋·萊文。Precog：以進球為條件的預測可視化多代理設定。IEEE/CVF 會議記錄
電腦視覺國際會議 (ICCV) 2019。2

[41] Christian Rupprecht、Iro Laina、Robert DiPietro、Maximilian Baust、Federico Tombari、Nassir Navab 與 Gregory D. Baust 海格。在不確定的世界中學習：透過多種假設表示歧義。在訴訟程序中
IEEE 國際電腦視覺會議 (ICCV) , 2017年2月

[42] Shaoshuai Shi, Li Jiang, Dengxin Dai, and Bernt Schiele。具有全局意圖定位和局部運動細化的運動變壓器。神經資訊處理系統 (NeurIPS) 的進展 2022 年。7

[43] Tong Su, Xishun Wang, and Xiaodong Yang. Qml for argoverse 2 運動預測挑戰。arXiv 預印本
arXiv:2207.06553, 2022. 7

[44] 馬修·坦西克、普拉圖·斯里尼瓦桑、本·米爾登霍爾、薩拉·Fridovich-Keil、Nithin Raghavan、Utkarsh Singhal、Ravi Ramamoorthi、Jonathan Barron 和 Ren Ng。傅立葉特徵讓網路學習低維域中的高頻函數。神經資訊處理系統進展 (NeurIPS) , 2020 年。4

[45] Yichuan Charlie Tang and Ruslan Salakhutdinov. Multiple 期貨預測。神經資訊處理系統進展 (NeurIPS) , 2019 年。2019 年。2

[46] Balakrishnan Varadarajan、Ahmed Hefny、Avikalp Srivastava、Khaled S Refaat、Nigamaa Nayakanti、Andre Cornman、Kan Chen、Bertrand Douillard、Chi Pang Lam、Dragomir 安格洛夫等人。Multipath++：高效率資訊融合和用於行為預測的軌跡聚合。國際機器人與技術會議論文集
自動化 (ICRA) 2022. 1, 2, 3, 5, 7

[47] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、ukasz Kaiser 和 Illia 波洛蘇欣。您所需要的就是關注。神經進展
資訊處理系統 (NIPS) , 2017 年。2

[48] Mingkun Wang, Xinge Zhu, Changqian Yu, Wei Li, Yuxin Ma, Ruochun Jin, Xiaoguang Ren, Dongchun Ren, Mingxu 王、楊文靜。Ganet：球門區域網絡
運動預測。arXiv 預印本 arXiv:2209.09723, 2022。7

[49] 班傑明·威爾森·威廉·齊·Tanmay Agarwal·約翰·蘭伯特·Jagjeet Singh·Siddhesh Khandelwal·Bowen Pan·Rat-nesh Kumar·Andrew Hartnett·Jhony Kaesemodel Pontes·德瓦·拉馬南·彼得·卡爾和詹姆斯·海斯·阿爾戈宇宙2：用於自動駕駛感知和預測的下一代資料集。神經資訊處理系統資料集和基準追蹤論文集 (NeurIPS

資料集與基準) ·2021. 1, 2, 6

[50]葉茂盛·曹同義·陳其峰。Tpcn：用於運動預測的時態點雲網路。IEEE/CVF 電腦視覺會議論文集

模式識別 (CVPR) ·2021. 1, 2, 7

[51] Maosheng Ye, Jiamiao Xu, Xunnong Xu, Tongyi Cao, and 陳奇峰. DCMS：具有雙重一致性和多偽目標監督的運動預測。arXiv 預印本 arXiv:2204.05859, 2022. 7

[52] Cunjun Yu, Xiao Ma, Jiawei Ren, Haiyu Zhao, and Shuai 義·用於行人軌跡預測的時空圖變換器網路。在歐洲議會訴訟中

電腦視覺會議 (ECCV) ·2020. 2

[53] Wenyuan Zeng, Ming Liang, Renjie Liao, and Raquel Urtasun. Lanercnn: Distributed representations for graph-centric 運動預測。在 IEEE/RSJ 國際智慧機器人和系統會議 (IROS) 會議記錄中，
2021年2月3日

[54] Chen Zhang, Honglin Sun, Chen Chen, and Yandong Guo. 2022 年 argoverse2 挑戰賽技術報告運動預測任務。arXiv 預印本 arXiv:2206.07934, 2022. 7

[55] Hang Zhao, Jiyang Gao, Tian Lan, Chen Sun, Benjamin Sapp, Balakrishnan Varadarajan, Yue Shen, Yi Shen, Yun-ing Chai, Cordelia Schmid, 等。Tnt：目標驅動的軌跡預測。機器人會議論文集 學習 (CoRL), 2020. 2, 3

[56] Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Ke-jie Lu. Hivt: Hierarchical vector transformer for multi-agent 運動預測。在 IEEE/CVF 計算機視覺和模式識別會議 (CVPR) 會議記錄中，
2022年1·2·3·6·7 _ _ _