

1 Human Trajectory Prediction via Neural Social Physics

1.1 Task

- 将一段轨迹分割为20帧 ($T = 20$)
- 输入: 前8帧的位置 ($M = 8$)
- 输出: 后12帧的位置

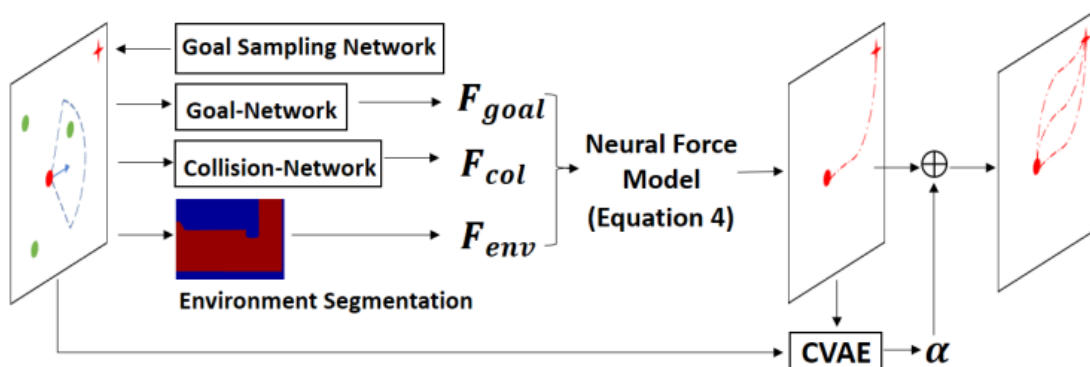
1.2 Formulation

- p_n^t : 第n个人在t时刻的位置
- \dot{p}_n^t : 第n个人在t时刻的速度
- $q_n^t = [p_n^t, \dot{p}_n^t]^T$: 第n个人在t时刻的状态
- Ω_n^t : 第n个人在t时刻邻居的集合
- E : 环境



邻居范围示意图

1.3 overview



$$\frac{dq}{dt}(t) = f_{\theta, \phi}(t, q(t), \Omega(t), q^T, E) + \alpha_{\phi}(t, q^{t:t-M}) \quad (1)$$

θ 、 ϕ 为可学习参数。

f 基于牛顿第二定律建模:

$$\ddot{p}(t) = F_{goal}(t, q^T, \dot{q}^t) + F_{col}(t, q^t, \Omega^t) + F_{env}(t, q^t, E) \quad (4)$$

α_ϕ 是扰动, 在以往的方法中用高斯噪声模拟, 但本文提出了学习这种扰动的方法。

状态更新:

$$q(t + \Delta t) \approx q(t) + \dot{q}(t)\Delta t = \begin{pmatrix} p(t) \\ \dot{p}(t) \end{pmatrix} + \Delta t \begin{pmatrix} \dot{p}(t) + \alpha(t, q^{t:t-M}) \\ \ddot{p}(t) \end{pmatrix} \quad (3)$$

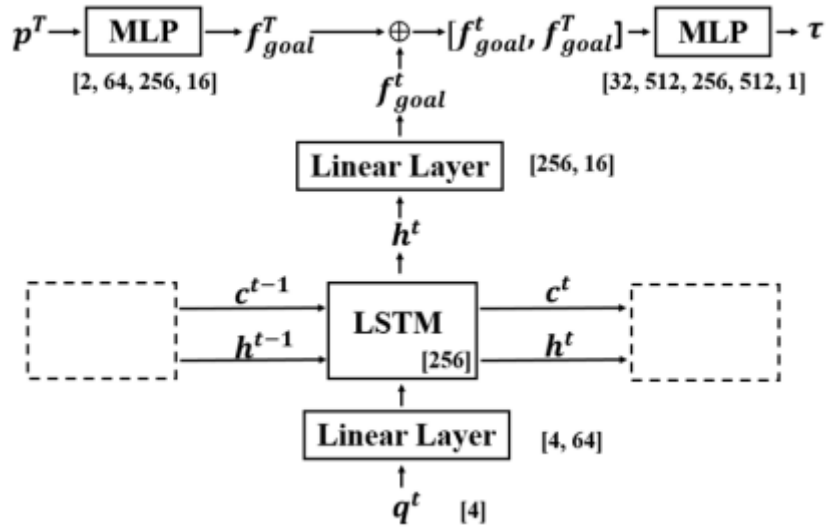
Goal Sampling Network是预训练好的模型, 用来设置终点。

1.4 Goal attraction

意义: 人朝着目标前进, 目标对人有吸引力

$$F_{goal} = \frac{1}{\tau} (v_{des}^t - \dot{p}^t) \text{ where } \tau = NN_{\phi_1}(q^t, p^T) \quad (5)$$

其中 $v_{des}^t = \frac{p^T - p^t}{(T-t)\Delta t}$.



1.5 Inter-agent Repulsion

意义: 个体会避免撞上自己邻居范围内的其他个体, 个体之间存在排斥力。

个体n与个体j间的排斥力:

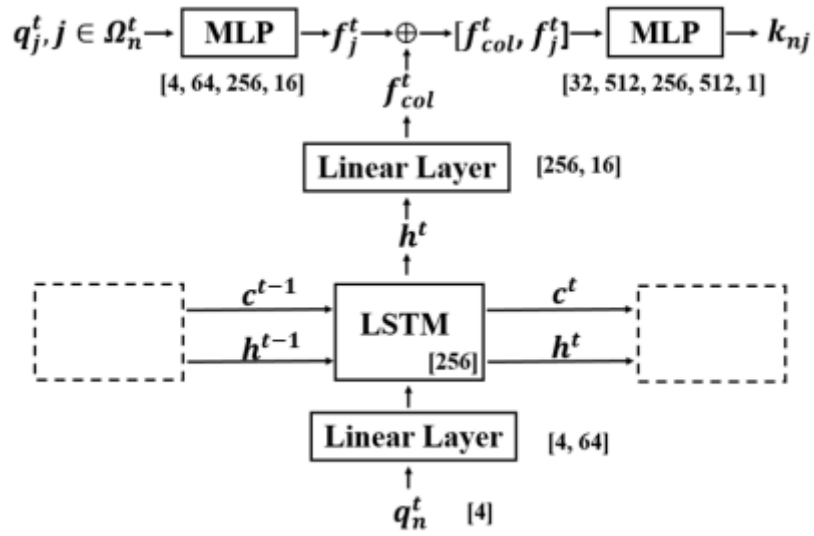
$$F_{col}^{nj} = -\nabla_{r_{nj}} \mathcal{U}_{nj}(\|r_{nj}\|), \text{ where } \mathcal{U}_{nj}(\|r_{nj}\|) = r_{col} k_{nj} e^{-\|r_{nj}\|/r_{col}} \quad (6)$$

个体n所受排斥力的总和 (假设个体的n的邻居范围内有m个个体):

$$F_{col}^n = \sum_{j=0}^m F_{col}^{nj}$$

$$k_{nj} = a * \text{sigmoid}(NN_{\phi_2}(q_n^t, q_{j, j \in \Omega_n^t}^t)) + b.$$

a, b 是超参数, 确保k有效。



1.6 Environment Repulsion

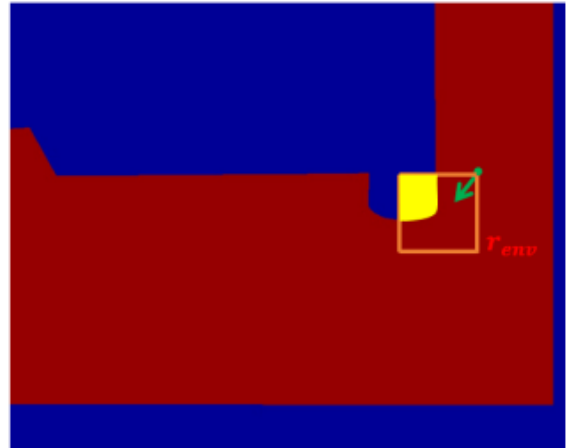
$$F_{env} = \frac{k_{env}}{\|p_n^t - p_{obs}\|} \left(\frac{p_n^t - p_{obs}}{\|p_n^t - p_{obs}\|} \right) \quad (7)$$

p_{obs} : 障碍物的位置 (视野内障碍区域的中心)

k_{env} 是可学习的参数 (无神经网络, 直接在 (7) 上反向传播)



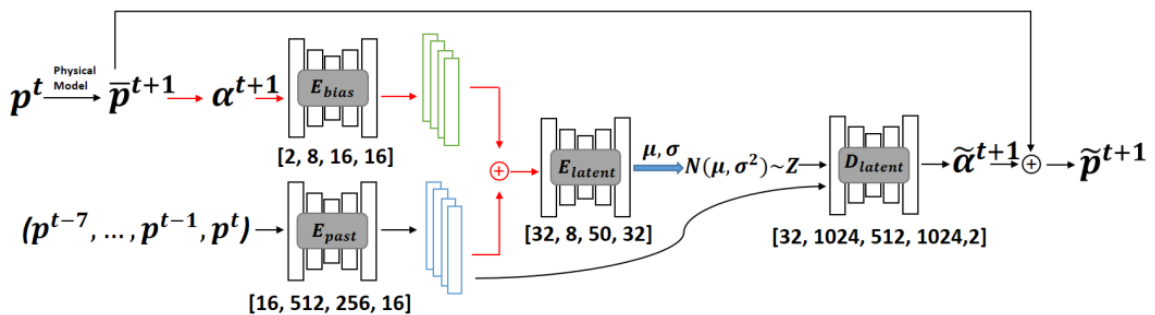
(b)



(c)

个体视野示意图 (红色为可行区域, 蓝色为障碍区域, 黄色为视野内的障碍区域)

1.7 Dynamics Stochasticity



故事: 为了更好地“constrain the learning”, 设计隐变量 Z 且假设它服从正态分布; 以往的方法直接假设扰动服从正态分布。

$$\alpha^{t+1} = \bar{p}^{t+1} - p^{t+1}.$$

红色路径仅在训练过程中使用。

训练过程中，正态分布的参数由 E_{latent} 提供；测试过程中，直接采取以下分布（ σ_{latent} 为超参数）

$$N(0, \sigma_{latent}I)$$

整体是CVAE：条件变分自动编码器

1.8 loss

$$L = l_{traj} + l_{cvae}$$

$$l_{traj} = \frac{1}{N(T-M)} \sum_{n=1}^N \sum_{t=M+1}^T \|p_n^t - \bar{p}_n^t\|_2^2$$

$$l_{cvae} = \frac{1}{N(T-M)} \sum_{n=1}^N \sum_{t=M+1}^T \{ \|\alpha_n^t - \tilde{\alpha}_n^t\|_2^2 + \lambda D_{KL}(Q(z|\alpha_n^t, p^{t:t-M}) || P(z|\alpha_n^t, p^{t:t-M})) \}$$
 (8)

一共 T 帧，前 M 帧为已知的输入，预测后面的 $T - M$ 帧。

D_{KL} ：相对熵，衡量两概率分布间的距离（ Q 为假设的正态分布， P 为真实分布）。

问题： P 从何而来？

参考<https://www.cnblogs.com/amazingter/p/14696251.html>， P 取 $N(0, 1)$ 。

1.9 Datasets

the Stanford Drone Dataset, ETH Hotel, ETH University, UCY University, Zara1, Zara2

1.10 Result

Average Displacement Error (ADE): 每一帧预测位置与实际位置的均方误差

Final Displacement Error (FDE): 预测的终点与实际终点的均方误差

Methods	Metrics	ETH	Hotel	UNIV	ZARA1	ZARA2	AVG	SDD
S-GAN [18]	ADE	0.81	0.72	0.60	0.34	0.42	0.58	27.23
	FDE	1.52	1.61	1.26	0.69	0.84	1.18	41.44
Sophie [49]	ADE	0.70	0.76	0.54	0.30	0.38	0.54	16.27
	FDE	1.43	1.67	1.24	0.63	0.78	1.15	29.38
CF-VAE [9]	ADE	N/A	N/A	N/A	N/A	N/A	N/A	12.60
	FDE	N/A	N/A	N/A	N/A	N/A	N/A	22.30
CGNS [29]	ADE	0.62	0.70	0.48	0.32	0.35	0.49	15.6
	FDE	1.40	0.93	1.22	0.59	0.71	0.97	28.2
NEXT [33]	ADE	0.73	0.30	0.60	0.38	0.31	0.46	N/A
	FDE	1.65	0.59	1.27	0.81	0.68	1.00	N/A
P2TIRL [14]	ADE	N/A	N/A	N/A	N/A	N/A	N/A	12.58
	FDE	N/A	N/A	N/A	N/A	N/A	N/A	22.07
SimAug [31]	ADE	N/A	N/A	N/A	N/A	N/A	N/A	10.27
	FDE	N/A	N/A	N/A	N/A	N/A	N/A	19.71
PECNet [38]	ADE	0.54	0.18	0.35	0.22	0.17	0.29	9.96
	FDE	0.87	0.24	0.60	0.39	0.30	0.48	15.88
Traj++ [50]	ADE	0.39	0.12	0.20	0.15	0.11	0.19	N/A
	FDE	0.83	0.21	0.44	0.33	0.25	0.41	N/A
Multiverse [32]	ADE	N/A	N/A	N/A	N/A	N/A	N/A	14.78
	FDE	N/A	N/A	N/A	N/A	N/A	N/A	27.09
Y-net [37]	ADE	0.28	0.10	0.24	0.17	0.13	0.18	7.85
	FDE	0.33	0.14	0.41	0.27	0.22	0.27	11.85
SIT [56]	ADE	0.38	0.11	0.20	0.16	0.12	0.19	N/A
	FDE	0.88	0.21	0.46	0.37	0.27	0.44	N/A
Social DualCVAE [16]	ADE	0.66	0.34	0.39	0.27	0.24	0.38	N/A
	FDE	1.18	0.61	0.74	0.48	0.42	0.69	N/A
CSCNet [71]	ADE	0.51	0.22	0.36	0.31	0.47	0.37	14.63
	FDE	1.05	0.42	0.81	0.68	1.02	0.79	26.91
NSP-SFM (Ours)	ADE	0.25	0.09	0.21	0.16	0.12	0.17	6.52
	FDE	0.24	0.13	0.38	0.27	0.20	0.24	10.61



红点：输入（即前M帧的已知路径）//绿点：预测轨迹//黑点：ground-truth

橙色箭头：终点吸引力//蓝色箭头：行人间排斥力//黑色箭头：障碍物排斥力

1.11 Generalization

在新场景中随机设置起点与终点，让测试对象沿着模型的预测轨迹运动。

由于此时没有ground truth，以对象的碰撞率作为评价标准（与障碍物或其他行人距离小于指定值时即发生碰撞）。

Methods	ETH	Hotel	UNIV	ZARA1	ZARA2	Avg	SDD
Y-net	0	0	1.51%	0.82%	1.31%	0.73%	0.47%
S-CSR	0	0	1.82%	0.41%	1.31%	0.71%	0.42%
NSP-SFM	0	0	1.48%	0	0.66%	0.43%	0.42%

1.12 Ablation Study

SDD	$F_{goal}(w/o)$	NSP-SFM(w/o)	NSP-SFM(w)
ADE	6.57	6.52	1.78
FDE	10.68	10.61	3.44

$F_{goal}(w/o)$: 终点吸引力

$NSP - SFM(w/o)$: 终点吸引力+行人间斥力+障碍物斥力

$NSP - SFM(w)$: 终点吸引力+行人间斥力+障碍物斥力+随机扰动