

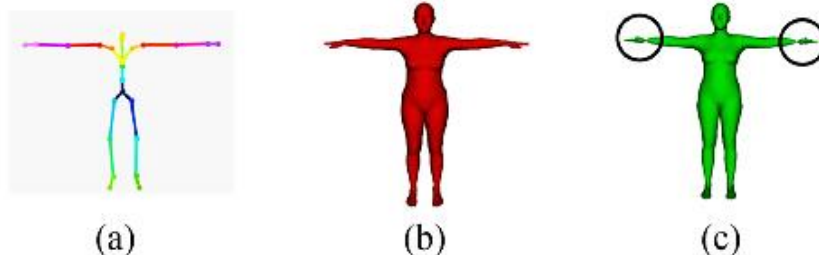
標題：bio-LSTM：猜出你的下一步

關鍵字：LSTM、SMPL、生物力學、損失函數



在轉彎的路口上，司機和行人正打著心理戰，誰先停、誰先動？如果沒有達成共識，在一聲巨響之後，便會留下無情的傷疤。近年來，自動車逐漸普及，「如何準確猜測對方的下一步」顯得更加重要，倘若 AI 能夠預先判斷行人往後的舉措和位置，便擁有較多空檔發展更完善的策略評估，除了增進人機互動的默契，也在速度和安全之間求得平衡，提升都市的交通效率。

## SMPL 模型



(圖一、2D 骨架與 3D 網格表示圖的差別)

事實上，人類姿態預測的研究不在少數。有些模型利用主要關節位置表示目標的姿勢，並以 2D 圖形輸出預測結果，減輕不少影像表示的麻煩；然而這個看似合理的設計，其實犧牲了不少細節，侷限了模型應用的潛力。舉例來說，圖

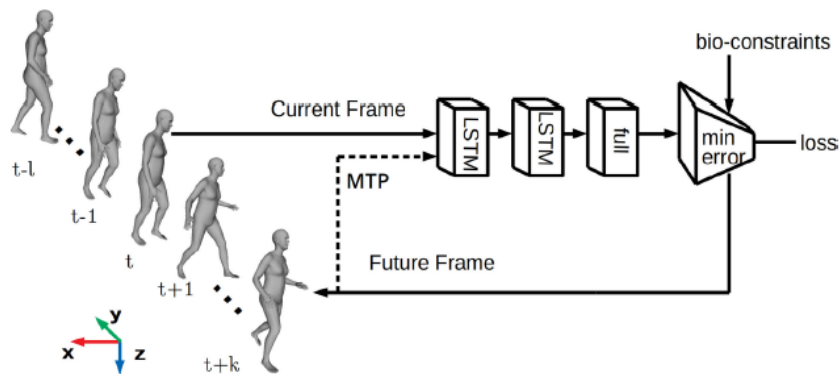
一中的(a)無法解釋目標的手心方向，雖說就目前的畫格而言似乎無傷大雅，但很可能影響接下來的猜測，做出完全截然不同判斷。

此外，有些模型甚至需要將目標置中定位，因此無法負擔多人預測，但這樣的限制將影響機器的判斷，尤其在攜來人往的十字路口上，路上的行人會同時移動，我們需要注意的不僅是視野中心的目標，而是整體交通環境的狀況。

為了讓研究更加逼真，bio-LSTM 就此誕生，除了採用擅長序列預測(sequence prediction)的長短期記憶網絡(即 LSTM)，還搭載 SMPL 模型(Skinned Multi-Person Linear Model)，藉由 85 個參數構成表面網格(mesh)，將先前的 2D 推廣為 3D 表示資料，其中包含 3 個紀錄人物與偵測系統距離的變換參數(translation parameter)、72 個表示關節位置和旋轉角度的姿勢參數(post parameter)以及 10 個代表人體輪廓的形狀參數(shape parameter)。

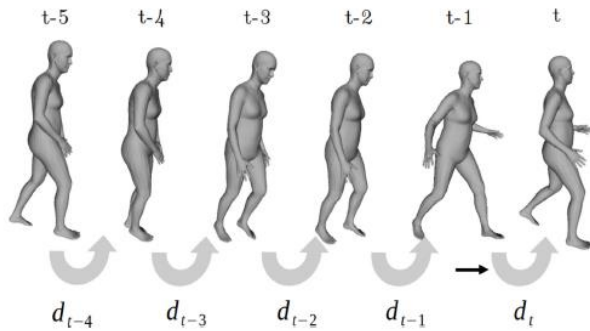
比起單純分析骨架關節位置，SMPL 在還原人物方面考慮了更多細節，有效降低預測的失誤，而變換參數也建立與目標在環境上的關係，推翻人物置中的假設，解決無法進行多人分析的窘境。

## 步態生物力學



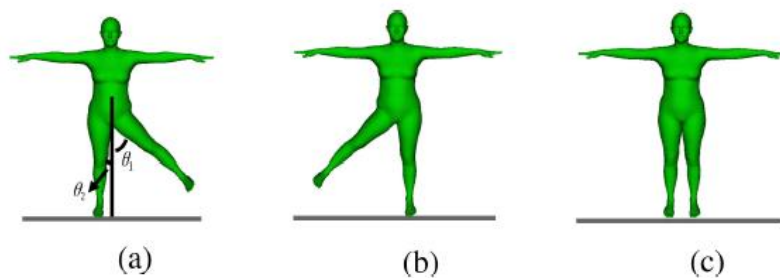
(圖二、bio-LSTM 結構)

bio-LSTM 的結構(如圖二所示)除了以兩層式 LSTM 循環神經網路為主幹、SMPL 參數來表示資料之外，還利用相關的領域基本知識(domain knowledge)掌握訓練方向，來修正預測運動時的不合理性。以下將介紹人類在穩定行進時的三種步態生物力學(gait biomechanics)，同時求出網路訓練的目標函數：



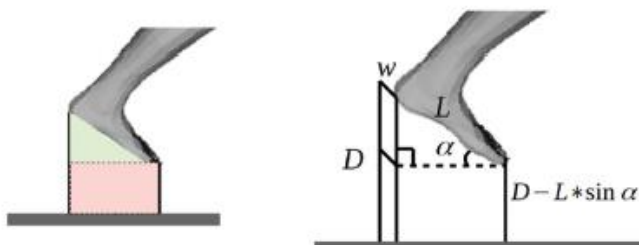
(圖三、 $dt$  表示時間  $t$  和時間  $t-1$  的資料差值)

首先，研究學者發現，人類在穩定行進時的步伐具有一定的規律，於是假設人腿會進行週期運動，其中相鄰畫格的腳速、方向和步長不會遽然大變，因此我們只需要推算畫格之間的差數即可(如圖三)。接著我們可以設立「週期損失函數 ( $Lc = |d_t - \hat{d}_t|$ )」，將預測的差數( $\hat{d}_t$ )和訓練資料的事實( $d_t$ )進行比對，調整出相對合理的運動狀態。



(圖四、 $\theta_1$ 和 $\theta_2$ 分別為左腿和右腿與鉛直線的夾角，(c)為平衡穩定狀態)

再者，人類在站立和行走時，身體應當呈現鏡像對稱(mirror symmetry)，維持左腳和右腳與鉛直線的夾角相等(如圖四)，肩膀與身體中線的夾角亦然，因此我們可以將「對稱損失函數」定為 $Ls = |\theta_{leg1} + \theta_{leg2}| + |\theta_{shoulder1} + \theta_{shoulder2}|$ ，以訓練推測結果的平衡感。



(圖五，腳底板被簡化為平面，方便計算離地體積)

最後，足夠的地面反作用力(Ground Reaction Force)是支撐身體的重要元素，而

力量主要由雙腳提供，導致人類在穩定行進時一定要與地面接觸，否則會跟踉蹌蹌。於是我們需要定義「GRF 損失函數( $Lg = wD(L \cos \alpha) -$

$\frac{1}{2}w(L \sin \alpha)(L \cos \alpha)$ )」，討論腳底離地的體積(如圖五)，考慮重力帶來的運動限制。

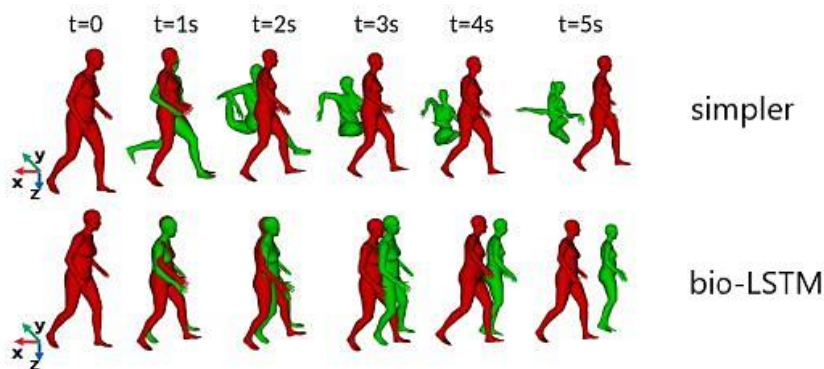
總歸以上三種生物力學知識，訓練預測人類穩定步行時的目標函數即為  $\min L = Lc + \lambda_1 Ls + \lambda_2 Lg$ 。而根據迴路測試的結果，其中的兩個正規化參數 (regularization parameter)  $\lambda_1$  和  $\lambda_2$  分別被定為 10 和 0.01，調整不同特徵對於結果的影響力。

## bio-LSTM 的前景



(圖六、下個畫格測試結果)

倘若今天 bio-LSTM 與自動車合作上路，它可以根據路人過去五個畫格的表現預測下一個畫格的狀況(人類的步伐周期大約五至六個畫格)，輸出結果(圖六綠人)和實況資料(圖六紅人)根本相去無幾，甚至當面對端著咖啡、扛著重物和滑著手機的行人時，也都能繳出相當卓越的成績，可見 bio-LSTM 的學習潛力果然不容小覷。



(圖七、MTP 預測結果)

那如果想要一次往後預測一段時間(數個畫格)呢？這種預測稱為 MTP(Multiple Timestep Prediction)，需要不斷將輸出餵回網路，成為新預測的參考資料(如圖二中的虛線)，但也可能因此導致誤差越滾越大，波及預測的參考價值。然而，bio-LSTM 的表現仍舊毫不遜色，和相對簡易的模型相比，結果明顯更加切實，合理還原人類穩定行進時大致的狀態(如圖七)。

雖然目前的 bio-LSTM 只考慮個體行為，但這都只是一個開端，大好前景還在不遠的未來等著。或許哪天我們真能透過不同領域知識的合作，打造完美的「行人友善路口」，將相關技術推廣研究機器人，還協助監督傷患的復健進度，點亮生活中一盞又一盞希望之光。

參考資料：

1. Xiaoxiao Du , Ram Vasudevan, and Matthew Johnson-Roberson, "[Bio-LSTM: A Biomechanically Inspired Recurrent Neural Network for 3D Pedestrian Pose and Gait Prediction](#)", 18 Feb 2019