中国科学: 数学 2021 年 第 51 卷 第 1 期
SCIENTIA SINICA Mathematica
论 文
《中国科学》杂志社
SCIENCE CHINA PRESS

融合神经网络与数值计算的人体逆向运动学求解

献给张景中、杨路教授 85 华诞

胡磊[1,2], 张子豪[1,2], 夏时洪[1,2*]

1. 中国科学院计算技术研究所前瞻研究实验室, 北京 100190;
2. 中国科学院大学计算机科学与技术学院, 北京 101408
E-mail: hulei19z@ict.ac.cn, zhangzihao@ict.ac.cn, xsh@ict.ac.cn

收稿日期: 2019-12-31; 接受日期: 2020-07-08; * 通信作者
国家自然科学基金 (批准号: 61772499) 和北京市自然科学基金 - 海淀原始创新联合基金重点项目 (批准号: L182052) 资助项目

摘要 人体逆向运动学问题是人体运动合成、人体运动捕获和理解的基本问题. 由于人体关节链式系统的复杂性, 人体逆向运动学方程往往存在多解或无解的情形. 传统的方法通常采用解析或数值迭代方法求解逆向运动学问题, 在给定足够多约束的情形下能够得到比较好的解, 但无法处理少量约束下生成自然的人体姿态问题. 近年来, 从大规模数据集中学习统计模型参数的思想被广泛运用, 求解人体逆向运动学的机器学习方法中经典工作—混合 Gauss 逆向运动求解模型 (Gaussian mixture model-inverse kinematics, GMM-IK) 就提出利用混合 Gauss 模型建模人体姿态数据分布, 并采用期望最大化方法求解参数. 随着深度学习技术的发展, 本文提出一种自编码神经网络与数值迭代融合的方法, 在给定少量约束的情形下依然能够得到自然的人体姿态, 相较于 GMM-IK 方法, 本文所提出的方法通过神经网络自动学习姿态分布, 省去了模型的假设和特征的设计, 且量化实验显示本文方法的关节坐标和角度重建误差相较于 GMM-IK 模型平均减少了 25% 和 39%. 在应用方面, 本文方法可处理光学运动捕获数据, 也可用于图像视频的人体姿态估计等领域.

关键词 逆向运动学 人体姿态构建 自编码神经网络

MSC (2010) 主题分类 34A34, 92B20, 53A17

1 引言

逆向运动学 (inverse kinematics, IK) 是利用运动学方程求解关节参数以满足位姿约束的问题 [1], 起初用来操纵 6 个自由度的机械手臂, 随后扩展到人体姿态的求解. 人体姿态的逆向运动学求解可以应用于计算机图形学领域的人体运动合成、计算机视觉领域的人体运动捕获与理解等方面, 例如, 无标注多风格的运动合成 [2]、基于动力学的运动合成与控制 [3] 以及基于单目深度相机的姿态估计 [4,5] 等. 机械手臂的逆向运动学问题由于自由变元个数少, 通过解析方法便能很好地解决. 但对于复杂的链式

英文引用格式: Hu L, Zhang Z H, Xia S H. MNN-IK: Mixing neural network and numerical IK for human posing (in Chinese). Sci Sin Math, 2021, 51: 1–16, doi: 10.1360/SSM-2019-0335

系统如人体, 由于约束的任意性以及约束个数与自由变元个数之间的关系导致约束方程组解的情形复杂多样, 且解的显式表达式难以给出. 因此, 人体逆向运动学求解问题一般都是基于数值迭代方法.

人体逆向运动学求解问题, 即给定一系列约束 (如部分关节点坐标等) 情形下, 求解满足约束条件的人体姿态 (各关节的旋转). 在此问题中, 给定基准姿态, 人体各关节的父子关系及在父节点局部坐标系下的坐标均为已知, 各关节相对于父关节的旋转角 θ 则是待求解的变量, 如图 1 所示.

父子关节点之间的坐标关系可表示为

$$X_a = X_f + R * O, \tag{1.1}$$

其中 $X_a = [x_a, y_a, z_a]^{\mathrm{T}}$ 表示关节点 a 在世界坐标系下的 3 维坐标, X_f 表示其父亲节点的世界坐标. R 为父节点在世界坐标系下的旋转矩阵, 与变量 θ 相关. $O = [o_x, o_y, o_z]^{\mathrm{T}}$ 表示在基准姿态下关节 a 在父节点的局部坐标系下的坐标.

当所有的关节点的旋转角度都已知时, 根据 (1.1), 从根节点出发, 根据关节链上节点的父子关系可以很容易求出所有关节点的 3 维坐标, 这即是正向运动学. 反过来, 知道所有或部分关节位置, 推导出各关节的旋转则困难得多, 通常情形下需要求解一个如 (1.2) 所示的非线性约束方程组:

$$\begin{cases} f_1(\theta) = c_1, \\ f_2(\theta) = c_2, \\ \quad \vdots \\ f_m(\theta) = c_m, \end{cases} \tag{1.2}$$

其中 θ 表示人体姿态的旋转变量, c_i 表示指定的约束, f_i 表示约束 c_i 与 θ 之间的函数关系.

在实际问题中, 约束包括但不局限于关节点的 3 维世界坐标, 其可以是用户指定的任意约束, 例如, 将 3 维人体模型投影至平面后的 2 维位置约束或是关节对之间的距离约束. 这些约束在增加直观性的同时也增加了求解的难度, 尤其是在约束的个数远小于待求解的未知数个数时, 例如, 图 2(a) 中红点即为相应关节 3 维坐标位置约束. 非线性方程组 (1.2) 存在着多解, 使得传统的数值迭代算法在该情形下难以求得自然的人体姿态, 如图 2(b) 所示.

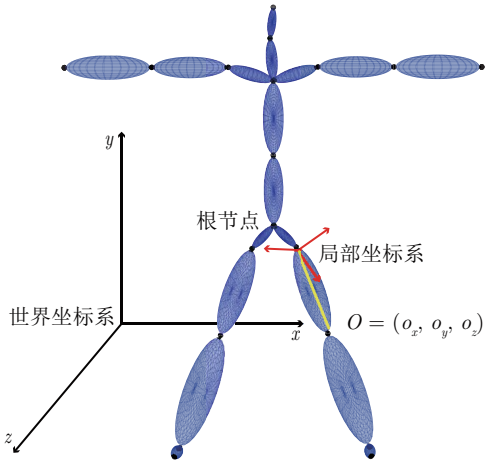

图 1　人体骨架示意图

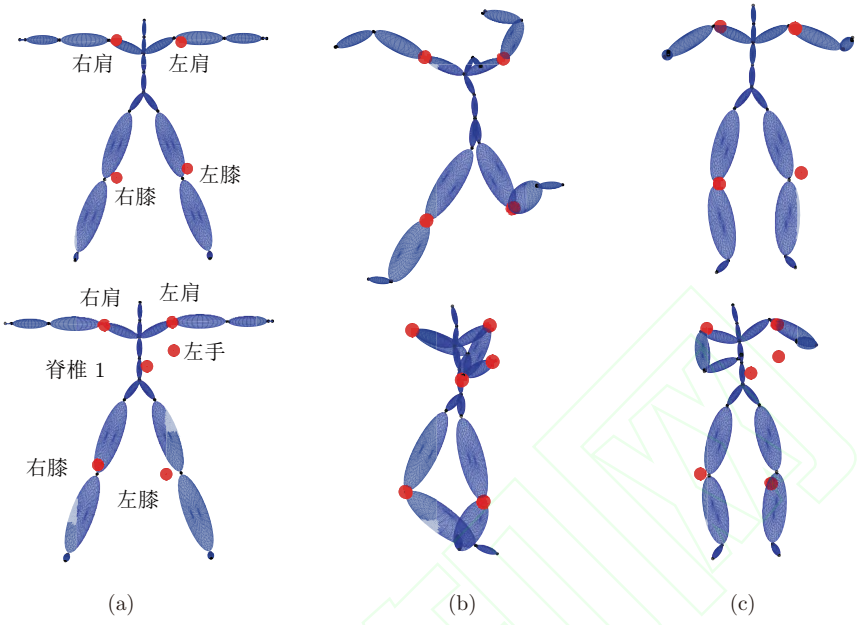

图 2 数值方法与神经网络求解逆向运动学. (a) 3 维关节坐标约束与人体基准姿态; (b) 纯数值迭代方法求解结果; (c) 神经网络方法求解结果

近年来深度学习发展迅速, 并已渗透到各个应用领域. 深度神经网络可以看作是非线性函数的拟合. 1989 年 Hornik 等 [6] 证明即便是只有一层隐藏层的神经网络, 使用任意的挤压函数 (squashing functions) 作为激活函数, 便可以拟合任意的 Borel 可测函数. 作为统计学习方法, 其思想主要是从大量的训练数据中自动提取特征, 并挖掘出其中的统计规律. 神经网络强大的拟合能力使其在 3 维人体重建和运动捕获等方面发挥着重要的作用. 但神经网络的方法仍然存在着弊端, 在人体逆向运动学问题中, 通过训练得到一个神经网络作为非线性函数来逼近方程组 (1.2) 真解的表达式, 但受到用户设定约束的多样性、网络结构以及训练数据规模的影响, 最终网络输出的解往往不能完全满足用户指定的约束, 如图 2(c) 所示.

因此, 本文研究了数值迭代计算方法和神经网络方法在少量约束下人体逆向运动学求解问题中的表现, 找到了将神经网络模型融入数值迭代计算的机制, 提出了融合神经网路与数值计算的人体逆向运动学求解模型 MNN-IK (mixing neural network and numerical IK for human posing), 使得求解结果达到了既数值准确、又视觉自然真实的效果, 进一步验证了该方法在光学运动捕获处理以及基于图像视频的人体姿态估计等实际应用中具有重要作用.

2 相关工作

求解逆向运动学问题的方法大致可以分为 3 类, 分别为解析方法、数值方法和数据驱动方法. 这 3 类方法均有各自的优点以及局限性.

2.1 解析方法

解析方法即是直接根据方程组 (1.2), 求得非线性方程组的闭式解 [7]. 这类方法的优点在于精确、直观明了且求解效率高. 在机器人领域, Manocha 和 Canny [8] 证明了解析方法能够很好地解决具有 6

个自由度的机械手臂操纵问题. 而在人体方面, Tolani 等[9] 结合解析和数值方法, 解决四肢关节的逆向运动学问题, 该方法能够实时地得到所有可能的解, 但仅限于所有关节旋转小于 7 个自由度的情形. 近年, Duits 等[10] 通过构造简单的三角函数来近似求解人手部的逆向运动学问题, 但仍然只能处理自由度较少的情形. 因此, 解析方法主要适用于简单系统的逆向运动学求解, 其缺陷在于对人体等复杂的链式系统, 解析方法很难给出关节角 θ 的显式表达式, 也难以处理方程组无解的情形.

2.2 数值方法

相对解析方法而言, 数值方法具有求解一般链式系统的形式且能够嵌入各种约束条件因而在求解人体逆向运动学方面更加地适用[11]. 逆向运动学的数值方法通过迭代来求解方程组 (1.2) 的近似解.

对于人体逆向运动学迭代算法的代表性研究按照时间顺序有基于 Jacobi 矩阵的迭代求解[12]、循环坐标下降法 (cyclic coordinate descent, CCD)[13] 求解以及 Levenberg-Marquardt (LM)[14] 方法等迭代算法.

Girard 和 Maciejewski[15] 最早提出利用 Jacobi 矩阵的伪逆来刻画关节坐标空间与关节角度空间的变化关系. 该算法每次令 θ 沿着 Jacobi 矩阵伪逆的方向更新迭代, 进而缩小坐标空间中关节位置与约束目标的距离. 该方法简单明了, 但实际求解 Jacobi 矩阵伪逆具有较高的时间复杂度, 且算法的收敛速度慢.

对 Jacobi 矩阵的迭代求解算法进行改进的是 CCD. CCD 是一种启发式的迭代优化求解算法, 每次仅优化 θ 中某一个变量, 而保持其他变量不变. 该方法具有较高的效率且对奇异点不敏感, 但由于 CCD 每次沿着空间中的坐标轴进行搜寻, 没有考虑人体关节链树形结构以及关节之间的关联, 因此当约束个数较少时, CCD 求解得到的人体姿态往往是不自然的.

与 CCD 一样, LM 法也同样能应用于逆向运动学方程的求解, LM 法结合了 Newton 法和梯度下降法的优势, 为了应对逆向运动学求解过程中的效率以及奇异值的问题, Sugihara[16] 对传统的 LM 法进行了改进, 提出了带阻尼因子的稳定 LM 求解方法, 该方法能够比较好地处理无解的情形, 且提升了求解的效率.

上述的工作在逆向运动学的迭代算法上均作出了巨大的贡献, 但均只是针对某一特定的约束或特殊情形. 而针对一般的人体逆向运动学求解, Zhao 和 Badler[17] 提出了普适的模型, 其将原方程组求解问题转化为非线性函数优化问题, 然后采用 LM 算法求解. 该模型能够适用于任意的链式模型, 且用户可以给定任意约束. 本文提出的 MNN-IK 即是融合了该优化模型的方法.

数值方法解决了当人体自由度较高时解析方法无能为力的问题. 但通过数值方法生成一个自然合理的人体姿态需要用户给定足够多的限制条件, 而实际应用往往只能得到少量的约束.

2.3 数据驱动方法

2.3.1 传统统计学习方法

近年来随着运动捕获设备的普及, 依赖数据的统计学习方法发展迅速, 利用已有的姿态数据作为先验信息合成自然合理的姿态成为目前研究的热点.

在角度空间中符合人体运动学的自然的姿态向量仅占少部分, 因此随着约束个数的减少, 数值求解结果的质量越加难以保证, 且极其依赖于迭代初值. 为此不少研究着眼于从相互关联的人体关节之间挖掘出相关规律, 而从高维度的角度空间中获取人体姿态先验信息的重要思想即是降维.

主成分分析 (principal component analysis, PCA) [18] 是经典的降维方法. 由于人体关节之间的相关性, 高维的角度空间必然存在着冗余的信息. Chai 和 Hodgins [19] 提出了利用局部 PCA 方法在低维空间中生成一个运动流形来合成运动姿态. 该方法在生成人体脚步运动方面较为有效, 但整个人体关节间的联系难以用线性模型表达.

为此, Li 等 [20] 提出了利用自组织映射 (self organizing map, SOM) [21] 的方法对姿态数据降维, SOM 是一种非线性无监督学习方法, 该工作通过学习特定运动序列的 Jacobi 矩阵来合成相似的人体运动. 该方法在小样本下效果较好, 且由于该方法通过学习得到 Jacobi 矩阵的伪逆, 因此求解效率高于一般的数值方法, 但其无法扩展到多约束以及多种运动风格的情形.

文献 [22] 提出利用 Gauss 过程隐变量模型 (Gaussian process latent variable model, GPLVM) [23] 学习特定风格的运动, 结合用户指定约束合成与训练动作相似的姿态. GPLVM 是低维非线性生成式模型, 其相较于 PCA 等线性方法具有更强大的生成能力.

以上统计学习方法共同的缺陷在于仅能从小样本中学习某一特定的运动风格, 针对大样本和多风格的情形, Wei 和 Chai [11] 利用线性方法先将高维的姿态向量降维, 然后将低维数据的分布由 70 个 Gauss 概率模型组成的 Gauss 混合模型 (Gaussian mixture model, GMM) [24] 来刻画, 最终通过最大期望 (expectation-maximization, EM) 算法 [25] 训练求解 GMM 的参数. 该方法能够从大量的数据中学习多种风格的运动姿态, 但随着训练数据规模增加, 其合成姿态的精度也随之下降, 量化实验证明本文的 MNN-IK 模型在姿态重建误差上优于 GMM-IK.

2.3.2 深度神经网络方法

神经网络既能拟合非线性函数, 也能学习得到数据集的先验. 前者主要用来做分类、回归等问题, 而后者在相关应用中的运用较为少见. Holden 等 [26] 于 2015 年提出了利用卷积神经网络 (convolutional neural networks, CNN) 学习低维人体运动流形空间, 该方法将时序的人体姿态数据映射到低维运动流形空间上, 进而应用于运动插值和填补缺失数据等, 但该方法并未利用到流形空间中运动数据的概率分布信息.

受到工作 [27] 的启发, 本文提出一种结合目前深度神经网络以及数值迭代求解的人体逆向运动学求解方法, 学习得到角度空间中表示姿态数据的先验分布, 以此来量化人体姿态的自然程度, 进而嵌入到原始的优化模型中来迭代求解. 相较于经典的数据驱动方法 GMM-IK 模型 [11], 本文的姿态先验是通过训练网络后利用重建误差直接得到, 从而省略了对概率模型以及相关参数的假设, 实验证明该方法在求解质量方面优于 GMM-IK 模型以及以往的方法.

3 MNN-IK 求解方法

3.1 MNN-IK 模型

由运动学方程组 (1.2), 可以发现其中方程仅包括用户指定的约束, 而缺少人体姿态的先验信息. 因此本文方法的主要思想很简单, 即引入一个判决函数 $g(\theta)$, 其数值大小仅与解 θ 本身有关, 反映的是解 θ 本身属于自然人体姿态的程度. 在求解过程中, 在满足用户约束的同时最小化判决函数, 以此保证合成的是自然的人体姿态. 为了方便后续的迭代求解, 函数 $g(\theta)$ 应当满足非负、可导并且 θ 表示姿态越自然则 $g(\theta)$ 的数值越小. 由此便将原问题和非线性方程组 (1.2) 建模为以下能量优化模型:

$$\min_{\theta} \frac{1}{2} E(\theta)^{\mathrm{T}} E(\theta), \tag{3.1}$$

其中 $E(\theta) = [c_1 - f_1(\theta), c_2 - f_2(\theta), \ldots, c_m - f_m(\theta), g(\theta)]^{\mathrm{T}}$.

残差矩阵 $E(\theta)$ 的前 m 项表示满足用户指定的约束条件, 而第 $m+1$ 项则保证该解表示的是属于人体的自然姿态. 因此关键部分是如何得到判决函数 $g(\theta)$, 为此将训练一个自编码网络 (autoencoder), 并且以自编码器对姿态的重建误差作为衡量解隶属于人体自然程度的数值指标.

受到 Yegnanarayana 和 Kishore [27] 工作的启发, 本文利用自编码网络来学习人体姿态先验, 再通过数值迭代求解优化问题 (3.1). 构造 $g(\theta)$ 的主要思想是利用大量的人体姿态数据来训练一个编码器和一个解码器, 见图 3. 由于人体在运动过程中各个关节之间存在着一定的联系, 因此本文希望通过降维来去除关节旋转角之间的冗余信息. 对于任意的输入, 网络首先将会通过编码器进行压缩降维, 得到低维的特征向量, 再通过解码器还原出输入. 该网络的目标是使得输入的姿态与输出的姿态尽可能地接近, 即最小化重建误差 $\|\theta - \hat{\theta}\|$, 因此本文训练所用的损失函数为

$$\mathcal{L} = \|\theta - \hat{\theta}\|^2 + cL_2, \tag{3.2}$$

其中 L_2 为正则化项 (网络权向量 W 中各元素的平方和再求平方根), 其目的是为了防止模型过拟合; c 为人为设置的超参数.

由于在训练时本文仅输入自然的人体的姿态 θ, 因此网络训练完成之后, 编码器与解码器便学会了编码与解码自然的人体姿态向量, 而对于其他非人体姿态的数据则不敏感, 从而重建误差也较大. 重建误差 $\|\theta - \hat{\theta}\|$ 则可以作为本文刻画人体姿态自然程度的指标来引导数值求解的方向. 因此可得到 $g(\theta)$ 的表达式为

$$g(\theta) = \|\theta - \hat{\theta}\|^2. \tag{3.3}$$

本文自编码网络的编码器与解码器均由 4 层全连接层构成, 具体网络结构及训练算法见附录 A: 自编码网络的训练. 由于网络中的计算均为矩阵乘法且修正线性单元 (rectified linear unit, ReLU) [28] 激活函数也为可导函数, 因此, $g(\theta)$ 可导且满足非负以及 $g(\theta)$ 的数值越小表示 θ 所表示的姿态越自然的要求.

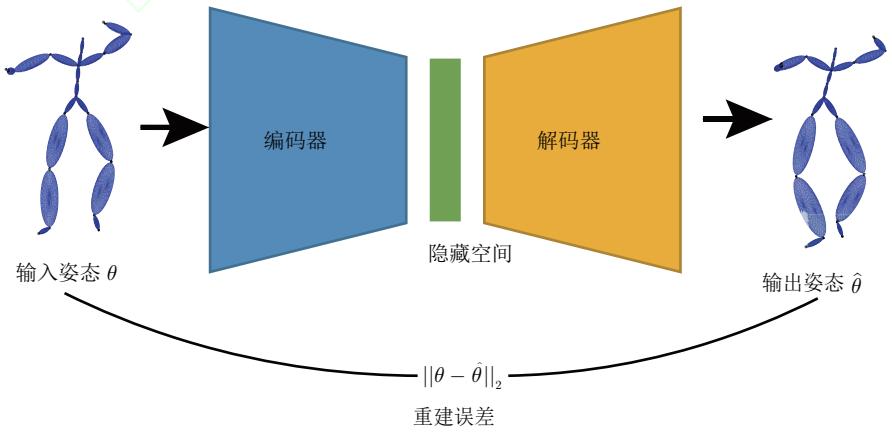

图 3 自编码器 (将高维空间中的人体姿态经过编码器得到在低维空间中的特征表示, 再通过解码器进行姿态的还原)

3.2　MNN-IK 算法

由训练好的自编码网络, 本文便可得到判决函数 $g(\theta)$, 从而可写出残差矩阵 $E(\theta)$, 接着通过 LM 法进行迭代求解, 即可求解得到满足约束的自然人体姿态. 参照 Lourakis [29] 的 LM 法实现, MNN-IK 求解算法的伪代码如算法 1, 其中, 参数 θ_0 表示迭代初值, 本文中取值为 0 向量; 参数 μ 调节 LM 公式中对角矩阵元素的大小, 反映的是矩阵 A 与单位矩阵 I 的比例关系; 参数 τ 则用来初始化 μ, 算法中 τ 的大小根据实验经验获得; 参数 v 表示算法内部循环失败之后对 μ 的调节, 即若算法没有找到正确的下降方向, 则将 μ 增大一倍从而加大二阶导数对迭代方向的影响. 该算法的收敛性分析参见附录 B.

算法 1　MNN-IK 模型求解

Require: 用户指定约束 c_i, 训练好的自编码器.
Ensure: 生成的姿态 θ.
初始化解 $\theta \leftarrow \theta_0$, 迭代步数 $k=0$, 最大迭代步数 $k_{\max}$, $v=2$, $\tau = 1\mathrm{e}-3$.
$\forall i$, 通过正向运动学方程计算出对应的 $f_i(\theta)$.
计算 $c_i - f_i(\theta)$, 以及 $g(\theta)$ 数值, 得到残差矩阵.
由 θ 求得 Jacobi 矩阵 J.
令 $A = J^{\mathrm{T}}J$, $g = J^{\mathrm{T}}E(\theta)$.
$\mu = \tau * \max_{i=1,\ldots,m+1}(A_{ii})$
while $E(\theta) > \epsilon_1$ 且 $k < k_{\max}$ **do**
  $k = k + 1$
  while 1 **do**
    计算 $\delta = (A + \mu I)^{-1} g$
    if $\|\delta\| \leqslant \epsilon_2$ **then**
      算法迭代结束
    else
      $\theta_{\mathrm{new}} = \theta + \delta$
      $\rho = (\|E(\theta)\|^2 - \|E(\theta_{\mathrm{new}})\|^2)/(\delta^{\mathrm{T}}(\mu\delta + g))$
      if $\rho > 0$ **then**
        $\theta = \theta_{\mathrm{new}}$
        根据新 θ 重新计算 $A = J^{\mathrm{T}}J$, $g = J^{\mathrm{T}}\epsilon$
        $\mu = \mu * \max(\frac{1}{3}, 1 - (2\rho - 1)^3)$, $v = 2$
      else
        $\mu = \mu * v$, $v = v * 2$
      end if
    end if
  end while
end while

4　实验

本节将首先介绍本文实验所用的数据集及数据预处理的方法, 第 4.2 小节将 MNN-IK 模型与传统的统计学习方法进行量化对比, 分别从关节坐标重建误差和关节旋转重建误差两个指标来证明本文方法的优越性, 同时将可视化一个迭代求解的过程来说明求解的高效性. 接着在第 4.3 小节展示对 MNN-IK 模型的定性研究结果, 对于每组姿态本文限定不同关节个数的 3 维坐标作为用户定义的约束, 然后利用 MNN-IK 模型进行逆向运动学求解. 定性研究结果说明给定不同的约束个数, MNN-IK 模型均能够合成出自然且符合约束的人体姿态.

4.1 实验数据集及预处理

本文采用的是 CMU-Mocap 数据集 [30], 数据包括日常生活的各类动作, 包含站立、行走、下蹲、跑步、跳舞和踢球等运动, 共计 500 余万帧. 等间隔采样将其中五分之三分为训练数据, 而五分之二分为测试数据. 去除了原始骨架中没有旋转量的 7 个末端关节点, 最终得到 31 个关节点的人体骨架. 给定 n 帧的人体姿态, 将姿态数据表示为一个 $n*(3*31)$ 的矩阵. 对于每个人体关节, 其旋转均由 Euler 角表示, 且旋转顺序为 Z、Y 和 X.

人体姿态的关节旋转变量是一个相对量, 其总是相对于一个基准姿态而言的 (如 T-pose 或 A-pose 等). 由于数据集中不同文件的基准姿态不同, 因此, 首先需要对所有数据的基准姿态进行一致化处理, 本文中将统一化为 A-pose. 对于深度学习而言, 为了让神经网络学习到姿态的统计规律, 需要将所有训练数据集中在同一局部坐标系而非整个 3 维空间, 因此根据根节点的位置将空间中的所有姿态都平移到 3 维世界坐标的原点处. 同时, 消除姿态在垂直坐标轴上的旋转, 即认定相对关节位置相同的姿态为同一姿态.

4.2 定量研究结果

采用两个评价指标来评估 MNN-IK 模型, 分别为关节点的平均坐标重建误差和关节平均旋转重建误差. 从传统机器学习方法中选择了几类典型的方法与本文提出的 MNN-IK 模型进行对比, 它们分别是线性降维方法概率 PCA [31]、非线性方法混合 Gauss 模型 [11] 和纯粹数值方法 [17]. 将这些方法分别在 CMU-Mocap 的测试数据集上进行测试, 将重建出的姿态与数据集中对应的真实姿态进行对比, 计算得到关节平均坐标重建误差和平均旋转重建误差. 由于当限制条件足够多时纯粹的数值方法便可以获得很好的重建结果, 姿态先验信息便失去了意义, 因此, 选择在少量的约束条件下 (在限定人体 6、8、10 和 15 个关节点的 3 维坐标下分别进行了实验) 比较各方法的重建优劣, 其中受到约束的关节是随机选定的. 从图 4 的平均重建误差水平来看, 本文所提出的方法具有较高的准确度, 且相较于 GMM-IK 模型, 关节坐标重建误差平均减少了 25%, 角度重建误差平均减少了 39% (4 种不同约束关节个数下误差下降率的平均值), 从二者的方差来看本文方法也具有更好的稳定性.

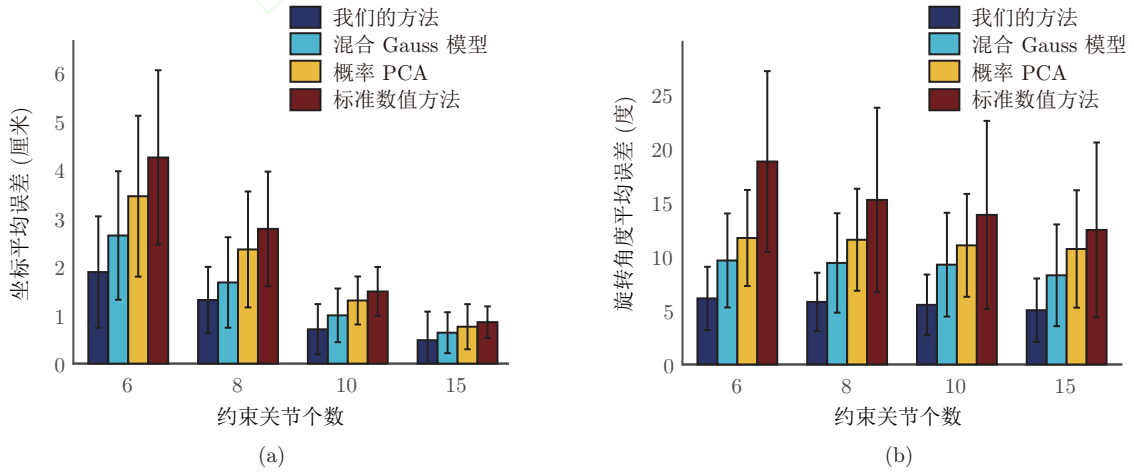

图 4 与其他方法的对比. (a) 在关节坐标上的重建对比; (b) 在关节旋转角度上的重建对比

4.2.1 迭代求解过程可视化

图 5(a) 显示了本文算法迭代过程中总误差、重建误差和判决函数值的下降过程. 图 5(b) 为相应的姿态迭代优化可视化结果, 其中红色的点表示设定的关节 3 维坐标约束. 求解过程说明了本文方法的有效性和高效性.

4.3 定性研究结果

该部分给定不同的 3 维坐标点个数来重建人体 3 维姿态, 用以说明本文方法在给定不同约束个数下均能生成合理的人体姿态. 人体姿态真值选自 CMU-Mocap 测试集的人体姿态, 给定部分人体约束如图 6(a) 和 6(c) 所示, 其中红点表示指定的关节坐标约束, 图 6(b) 和 6(d) 为本文的生成结果. 从

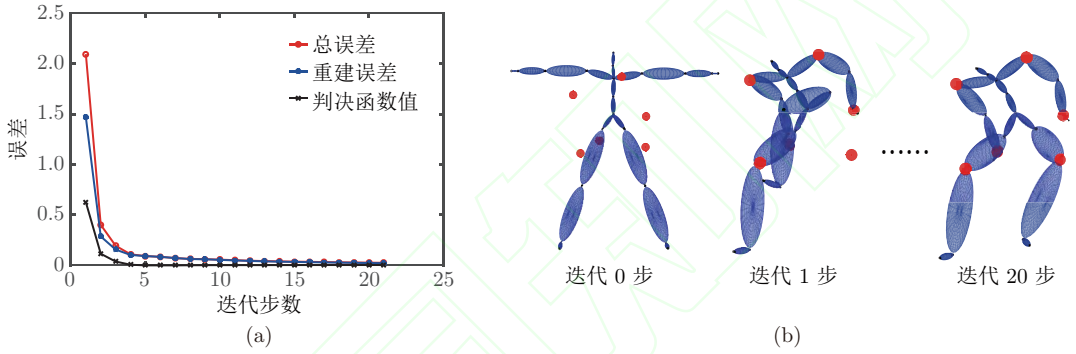

图 5 迭代求解过程可视化. (a) 迭代求解误差曲线图; (b) 迭代求解姿态可视化

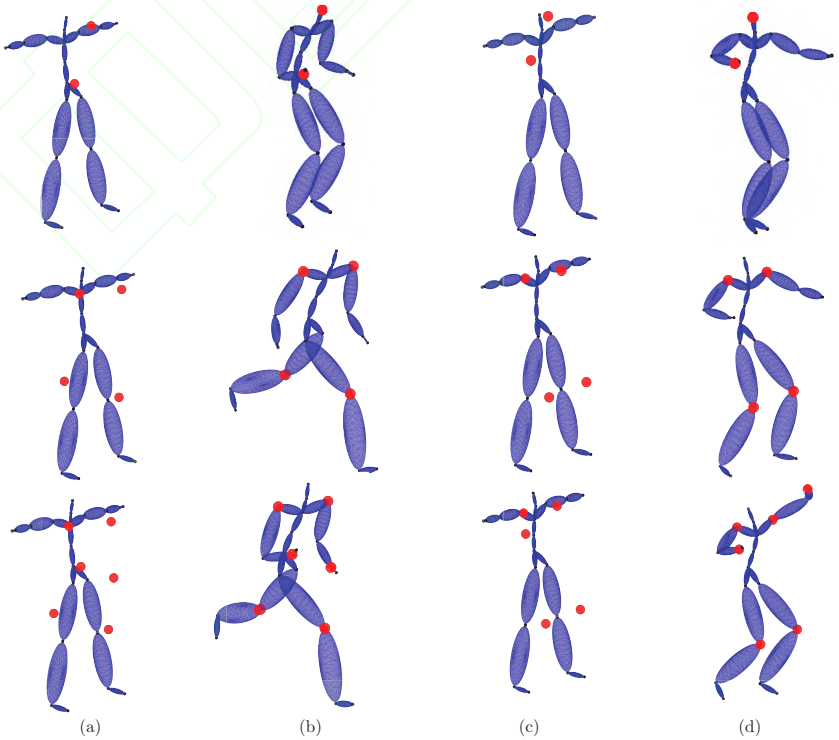

图 6 不同约束个数下 3 维人体姿态生成. (a) 和 (c) 用户约束及初始姿态; (b) 和 (d) 合成结果

图中可以观察到, 在给定不同约束个数下, 本文方法均能够生成合理且满足用户指定约束限制的姿态, 体现了其具有普遍适用性.

5 应用

MNN-IK 模型在计算机视觉的相关任务上均具有一定应用价值, 包括光学运动捕获数据处理和基于图像视频的人体姿态估计.

5.1 光学运动捕获数据处理

光学运动捕获系统因稳定、准确的特点依然占据着运动捕获的主要市场, 其原理是, 先捕获标志点 3 维坐标, 然后依据标志点相对各关节的位置, 通过求解人体运动学约束方程计算人体的姿态 θ. 但由于环境和设备的原因, 标志点的坐标位置有时会存在着误差, 甚至出现标记点缺失的情形, 这时便需要人工进行补点或修点, 十分地费时费力. 利用本文方法, 在求解人体姿态 θ 的时候引入判决函数 $g(\theta)$, 即可根据人体姿态的先验信息还原出合理的人体姿态, 数学模型为

$$\min \frac{1}{2} E_{\text{cap}}(\theta)^{\mathrm{T}} E_{\text{cap}}(\theta), \tag{5.1}$$

其中

$$E_{\text{cap}}(\theta) = [c_1 - f_1(\theta), c_2 - f_2(\theta), \ldots, g(\theta)]^{\mathrm{T}}$$

与上文一样表示残差矩阵, 只是 c_i 表示标志点的 3 维坐标, f_i 表示标志点坐标与姿态 θ 之间的函数关系.

图 7(a) 展示了在捕捉过程中右手标记点严重缺失的情形, 图 7(b) 为通过本文方法重建出的结果, 可以看到在右手仅捕获到一个标记点情形下仍然能得到较好的结果. 除了能应对标记点的缺失, 同样对标记点的输入具有降噪处理的功能. 图 7(c) 展示标记点受到噪声影响发生偏移的情形, 可以看到尤其是双手交叉的部位, 因为标记点较为集中, 人为修点降噪处理费时费力, 而通过本文方法能够较好地处理该问题, 图 7(d) 为重建的结果.

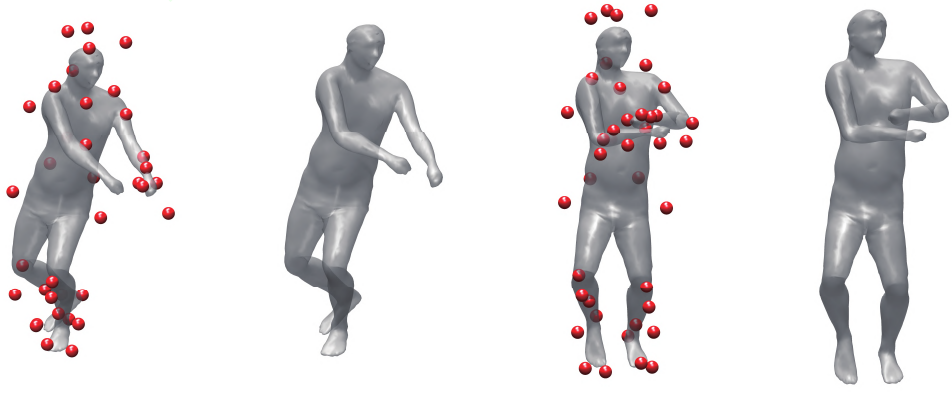

图 7 MNN-IK 在光学运动捕获上的应用. 红色球体代表捕获运动时的标记点. (a) 右手标记点缺失; (b) 重建结果; (c) 带有噪声的标记点; (d) 重建结果

5.2 基于图像视频的人体姿态估计

基于单目 RGB 图像视频的 3 维姿态估计也是目前的热点问题, 且在如智能驾驶应用等方面有着重要的作用. 由于 RGB 相机在运动捕获当中无法获得人体的深度值, 因此基于单目 RGB 相机的人体姿态估计成为研究的难点. 目前普遍做法是通过神经网络自动推测 3 维人体关节的 3 维坐标点, 但由于自遮挡等问题导致许多姿态估计得并不是很准确. MNN-IK 模型在 HMR (human mesh recovery) 模型 [32] 的基础上进行了优化, 将原本无法满足 2 维约束的错误姿态进行了修正. 首先利用 HMR 模型来预测出相机的参数 (简化为正交投影) s、x_0 和 y_0, 则可得 3 维空间到 2 维平面的投影公式为

$$(x_{2D}, y_{2D}) = s(x_{3D} - x_0, y_{3D} - y_0), \tag{5.2}$$

其中 s 表示缩放因子, x_0 和 y_0 分别表示由 3 维坐标到图像平面坐标的偏移. 将投影函数简记为 proj, 便可以将原应用问题建模为最小化人体关节投影误差及判决函数 $g(\theta)$ 的问题, 如

$$\frac{1}{2}E_{2D}(\theta)^{\mathrm{T}}E_{2D}(\theta), \tag{5.3}$$

其中残差矩阵 $E_{2D} = [C_{2D} - \mathrm{proj}(f(\theta)), g(\theta)]^{\mathrm{T}}$.

从图 8 可以看到, 本文方法能得到更合理和自然的结果, 其中 RGB 图像取自于 Ms CoCo [33] 验证数据集.

(a) (b) (c) (d)

图 8 各类型方法在 RGB 姿态估计上的对比. (a) 原始 RGB 图像以及关键点; (b) HMR 神经网络模型结果; (c) 数值 IK 的结果; (d) MNN-IK 模型优化后的结果

6 结论

本文提出了一种数值计算与深度神经网络融合的人体逆向运动学求解方法. 给定一系列任意的用户约束该方法能够合成出满足约束且合理的人体姿态. 实验证明本文的混合方法能够弥补目前深度学习方法的不足, 且在人体姿态重建方面, MNN-IK 模型相较于经典工作—GMM-IK 模型在关节坐标重建误差上平均减少了 25%, 在角度误差上平均减少了 39%.

值得关注的是本文方法在人体运动分析与理解领域具有良好的应用潜力. 由于其能够在给定少量限制条件时得到较好的结果, 因此, MNN-IK 模型在运动捕获中能够减少人工补点和修点的工作量, 同时, 也为基于图像视频的姿态估计解决自遮挡和姿态歧义性难题提供了一个解决思路.

不过, 在本文的求解模型中仍然存在着一些局限有待进一步的研究解决.

(1) 由于自编码的重建误差得到的判决函数 $g(\theta)$ 结构的复杂性, 在优化过程中容易陷入局部最优点, 导致算法不能收敛到较好的值.

(2) 目前通过训练自编码器构造判决函数 $g(\theta)$ 仍然依赖于大量的人体姿态样本, 如何在少量样本上进行训练学习是未来研究的方向.

对于以上存在的问题, 存在着几种可能的改进方法来进一步优化本文的模型. 针对算法求解过程中存在着落入局部最优的可能, 可以设计一个初值预测网络, 首先将用户的特定输入转化为关节点的约束, 然后由一个深度神经网络预测一个良好的初值, 这样能够进一步地提高模型的求解效率和稳定性. 针对在小规模样本上训练模型的问题, 可以考虑引入对抗生成网络 (generative adversarial nets, GAN) 等方式来进一步解决.

致谢 感谢审稿人提出的建设性修改建议.

参考文献

1 Paul R P. Robot Manipulators: Mathematics, Programming, and Control: The Computer Control of Robot Manipulators. Cambridge: MIT Press, 1981

2 Xia S, Wang C, Chai J, et al. Realtime style transfer for unlabeled heterogeneous human motion. ACM Trans Graph, 2015, 34: 119

3 Lv X, Chai J, Xia S. Data-driven inverse dynamics for human motion. ACM Trans Graph, 2016, 35: 163

4 Xia S, Zhang Z, Su L. Cascaded 3D full-body pose regression from single depth image at 100 FPS. In: Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Piscataway: IEEE, 2018, 431–438

5 Zhang Z H, Hu L, Deng X M, et al. Weakly supervised adversarial learning for 3D human pose estimation from point clouds. IEEE Trans Visual Comput Graphics, 2020, 26: 1851–1859

6 Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks, 1989, 2: 359–366

7 Korein J U, Badler N I. Techniques for generating the goal-directed motion of articulated structures. IEEE Comput Graph Appl, 1982, 2: 71–81

8 Manocha D, Canny J F. Efficient inverse kinematics for general 6R manipulators. IEEE Trans Robot Automat, 1994, 10: 648–657

9 Tolani D, Goswami A, Badler N I. Real-time inverse kinematics techniques for anthropomorphic limbs. Graph Model, 2000, 62: 353–388

10 Duits R, Egges A, van der Stappen A F. A closed-form solution for human finger positioning. In: Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games. New York: Association for Computing Machinery, 2015, 73–78

11 Wei X, Chai J. Intuitive interactive human-character posing with millions of example poses. IEEE Comput Graph Appl, 2009, 31: 78–88

12 Wolovich, William A, Elliott H. A computational technique for inverse kinematics. In: Proceedings of the 23rd IEEE Conference on Decision and Control. Piscataway: IEEE, 1984, 1359–1363

13 Komura T, Kuroda A, Kudoh S, et al. An inverse kinematics method for 3D figures with motion data. In: Proceedings of the 2003 Computer Graphics International Conference. Los Alamitos: IEEE, 2003, 266–271
14 Levenberg K. A method for the solution of certain non-linear problems in least squares. Quart Appl Math, 1944, 2: 164–168
15 Girard M, Maciejewski A A. Computational modeling for the computer animation of legged figures. ACM SIGGRAPH Comput Graph, 1985, 19: 263–270
16 Sugihara T. Solvability-unconcerned inverse kinematics based on Levenberg-Marquardt method with robust damping. In: Proceedings of the 9th IEEE-RAS International Conference on Humanoid Robots. Piscataway: IEEE, 2009, 555–560
17 Zhao J, Badler N I. Inverse kinematics positioning using nonlinear programming for highly articulated figures. ACM Trans Graph, 1994, 13: 313–336
18 Pearson K. On lines and planes of closest fit to systems of points in space. Philos Mag, 1901, 2: 559–572
19 Chai J, Hodgins J K. Performance animation from low-dimensional control signals. ACM Trans Graph, 2005, 24: 686–696
20 Li C, Xia S, Wang Z. Pose synthesis using the inverse of Jacobian matrix learned from examples. In: Proceedings of the 2007 IEEE Virtual Reality Conference. Los Alamitos: IEEE, 2007, 99–106
21 Kohonen T. Self-organized formation of topologically correct feature maps. Biol Cybernet, 1982, 43: 59–69
22 Grochow K, Martin S L, Hertzmann A, et al. Style-based inverse kinematics. ACM Trans Graph, 2004, 23: 522–531
23 Lawrence N. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. J Mach Learn Res, 2005, 6: 1783–1816
24 Newcomb S. A generalized theory of the combination of observations so as to obtain the best result. Amer J Math, 1886, 8: 343–366
25 Dempster A P, Laird N M, Rubin D B. Discussion on the paper by Professor Dempster, Professor Laird and Dr Rubin. J R Stat Soc Ser B Stat Methodol, 1977, 39: 22–38
26 Holden D, Saito J, Komura T, et al. Learning motion manifolds with convolutional autoencoders. In: SIGGRAPH Asia 2015 Technical Briefs. New York: Association for Computing Machinery, 2015, 18
27 Yegnanarayana B, Kishore S P. AANN: An alternative to GMM for pattern recognition. Neural Networks, 2002, 15: 459–469
28 Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning. Madison: Omnipress, 2010, 807–814
29 Lourakis M I A. A brief description of the Levenberg-Marquardt algorithm implemented by levmar. Found Res Technol, 2005, 4: 1–6
30 Carnegie Mellon university motion capture database. http://mocap.cs.cmu.edu/
31 Chai J X, Hodgins J K. Performance animation from low-dimensional control signals. In: Proceedings of SIGGRAPH05: Special Interest Group on Computer Graphics and Interactive Techniques Conference. New York: Association for Computing Machinery, 2005, 686–696
32 Kanazawa A, Black M J, Jacobs D W, et al. End-to-end recovery of human shape and pose. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE, 2018, 7122–7131
33 Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context. In: Computer Vision—ECCV 2014. Cham: Springer, 2014, 740–755

附录 A 自编码网络的训练

附录 A.1 自编码神经网络的前向传播

本文的网络各层均为全连接, 即前一层各神经元与后一层的每一个神经元均具有联系, 如图 A1. 假设网络 L 层具有 m 个神经元, 而 $L+1$ 层具有 n 个神经元, 那么网络由 L 层向 $L+1$ 层的传播公式为

$$\sigma(W^L X^L + b^L) = X^{L+1}, \tag{A.1}$$

其中 W^L 为一个维度为 $m \times n$ 的权重矩阵; b^L 表示偏置量, 维度为 $n \times 1$; σ 表示激活函数, 且 σ 是

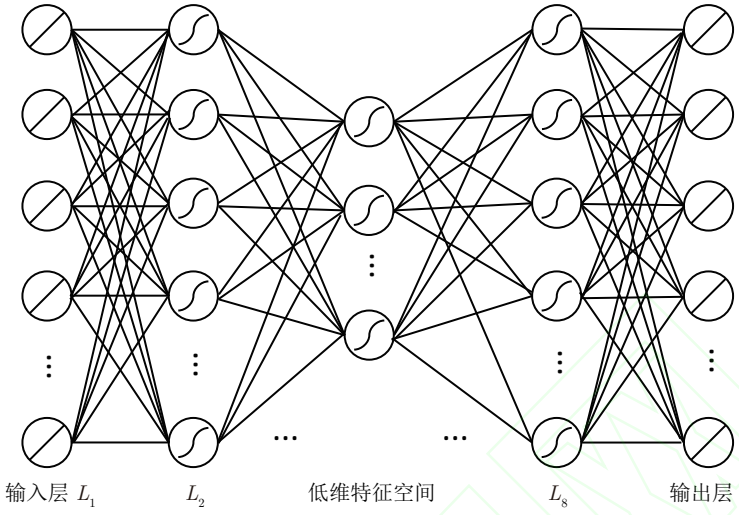

图 A1　自编码神经网络模型连接图

非线性的. (假使 σ 为线性函数, 则神经网络中的多层连接等价于一个单层网络, 这将会失去深度的意义.)

有了神经网络的前向传播公式, 给定一个姿态表示为 θ 的输入, 网络最终输出与输入维度相同的向量 $\hat{\theta}$. 根据本文的目标, 应当使得 θ 与 $\hat{\theta}$ 足够接近. 因此, 损失函数设计为

$$\mathcal{L} = \|\theta - \hat{\theta}\|^2 + cL_2, \tag{A.2}$$

其中 L_2 为正则化项 (网络权向量 W 中各元素的平方和再求平方根), 其目的是为了防止模型过拟合; c 为人为设置的超参数.

附录 A.2　自编码神经网络的反向传播

有了网络的前向传播过程和损失函数 (A.2), 仅需要设计算法求解得到网络中的参数使得损失函数最小化即可, 求解网络参数的反向传播算法如下. 已知一个多元函数的梯度方向是该函数在该点上升最快的方向, 因此, 每次迭代优化神经网络中的参数总是朝着负梯度方向进行优化, 如

$$\begin{aligned} W^L &= W^L - \alpha \frac{\partial \mathcal{L}}{\partial W^L}, \\ b^L &= b^L - \alpha \frac{\partial \mathcal{L}}{\partial b^L}, \end{aligned} \tag{A.3}$$

其中 α 表示学习率, 则关键步骤便是计算偏导数, 而计算网络各层参数偏导数的方法即是网络的反向传播算法.

用 z^{L+1} 表示网络经过前 L 层的前向计算但未被激活函数作用的结果, 用 a^{L+1} 表示激活后的结果, 用公式表达即为

$$z^{L+1} = W^L a^L + b^L, \quad a^L = \sigma(z^L). \tag{A.4}$$

利用求导的链式法则将 $\frac{\partial \mathcal{L}}{\partial W^L}$ 拆分得到

$$\frac{\partial \mathcal{L}}{\partial W^L} = \frac{\partial \mathcal{L}}{\partial z^{L+1}} \frac{\partial z^{L+1}}{\partial W^L}. \tag{A.5}$$

根据 (A.4) 可知, $\frac{\partial z^{L+1}}{\partial W^L} = a^L$, 令 $\frac{\partial \mathcal{L}}{\partial z^{L+1}} = \delta^{L+1}$, 则有

$$\frac{\partial \mathcal{L}}{\partial W^L} = \delta^{L+1} a^L. \tag{A.6}$$

对于 δ^{L+1} 与 δ^L 之间的关系, 有

$$z^{L+1} = W^L \sigma(z^L) + b^L, \tag{A.7}$$

所以, $\delta^L = (W^L)^{\mathrm{T}} \delta^{L+1} \sigma'(z^L)$, 因此依据后一层的偏导数 δ^{L+1}, 能够很容易地求出偏导数 δ^L, 再依据 (A.6) 便能求解出网络中参数 W 的偏导数, 而参数 b 偏导数的求法与之类似.

本文的模型在 Tensorflow 深度学习框架下实现, 网络激活函数为 ReLU 函数, 初始的学习率为 0.0001, 且每隔 2,000 步以 0.98 的衰减率衰减. 每次训练使用的批量大小 $N = 10{,}000$, 网络中各参数用均值为 0、方差为 0.01 的正态分布来进行初始化. 网络中间低维特征维度为 35, 各全连接层的神经元个数均为 1,024 个.

附录 B MNN-IK 算法收敛性分析

MNN-IK 模型求解算法收敛性与 LM 法收敛性相同. 根据用户指定约束条件, 将人体逆向运动学问题化为如 (3.1) 的优化问题.

为此, 可以先求得 (3.1) 的极值点, 设第 k 步的姿态为 $\theta^k = (\theta_1^k, \theta_2^k, \ldots, \theta_n^k)$, 将 Newton 法推广至多元函数的情形, 那么姿态 θ 的更新公式如

$$\theta^{k+1} = \theta^k - \frac{\nabla E(\theta^k)^{\mathrm{T}} E(\theta^k)}{\nabla(\nabla E(\theta^k)^{\mathrm{T}} E(\theta^k))}, \tag{B.1}$$

其中 $\nabla(\nabla E(\theta^k)^{\mathrm{T}} E(\theta^k)) = \nabla^2 E(\theta^k)^{\mathrm{T}} E(\theta^k) + \nabla E(\theta^k)^{\mathrm{T}} \nabla E(\theta^k)$. 将 $\nabla^2 E(\theta^k)$ 展开得

$$\nabla^2 E(\theta^k) = \sum_{i=1}^{n} \frac{\partial \nabla E(\theta^k)}{\partial \theta_i^k}. \tag{B.2}$$

因为 $\nabla^2 E(\theta^k)$ 的值一般远小于 $\nabla E(\theta^k)^{\mathrm{T}} \nabla E(\theta^k)$ 的值, 所以可以用 $\nabla E(\theta^k)^{\mathrm{T}} \nabla E(\theta^k)$ 的值来近似替代 $\nabla(\nabla E(\theta^k)^{\mathrm{T}} E(\theta^k))$. 令

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1^k} & \frac{\partial f_1}{\partial \theta_2^k} & \cdots & \frac{\partial f_1}{\partial \theta_n^k} \\ \frac{\partial f_2}{\partial \theta_1^k} & \frac{\partial f_2}{\partial \theta_2^k} & \cdots & \frac{\partial f_2}{\partial \theta_n^k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial \theta_1^k} & \frac{\partial f_m}{\partial \theta_2^k} & \cdots & \frac{\partial f_m}{\partial \theta_n^k} \\ \frac{\partial g}{\partial \theta_1^k} & \frac{\partial g}{\partial \theta_2^k} & \cdots & \frac{\partial g}{\partial \theta_n^k} \end{bmatrix}.$$

经过近似化后迭代公式变为

$$\theta^{k+1} = \theta^k + (J^{\mathrm{T}} J)^{-1} J^{\mathrm{T}} E(\theta^k). \tag{B.3}$$

将 $\nabla^2 E(\theta)$ 省略掉后, 因为 $J^{\mathrm{T}} J$ 不一定可逆, 为此引入单位矩阵, 迭代式变为

$$\theta^{k+1} = \theta^k + (J^{\mathrm{T}} J + \mu I)^{-1} J^{\mathrm{T}} E(\theta^k), \tag{B.4}$$

其中 μ 将根据 Jacobi 矩阵自动调节, 具体见算法 1. 由 Newton 法性质可知 MNN-IK 模型的 LM 求解算法具有二阶收敛性.

MNN-IK: Mixing neural network and numerical IK for human posing

Lei Hu, Zihao Zhang & Shihong Xia

Abstract The inverse kinematics of human figure is the basic problem of human motion synthesis, human motion capture and understanding. The system of kinematic constraint equations has many or no solutions to the inverse kinematics for human posing because the human system is very complex and its articulated representation has many degrees of freedom. Traditional methods usually use analytical or numerical iterative methods to solve the inverse kinematics problem. They can obtain good results with sufficient constraints. However, it is difficult to find natural human pose only given a small number of constraints. In recent years, the idea of learning statistical models from large-scale data sets has been widely used. Among the machine learning methods for solving inverse kinematics of human body, the classical work GMM-IK proposes to use Gaussian mixture model to construct the human posture data distribution, and uses the expectation maximization method to solve the parameters. This paper presents a method combining neural networks and numerical inverse kinematics for human posing. It can synthesize natural human pose with a small number of constraints. Extensive quantitative experiments show that the joint coordinates and angle reconstruction errors of our method are reduced by an average of 25% and 39%, respectively, compared with the state of the art, i.e., Gaussian mixture model. Our method can be used to deal with optical motion capture data, and estimate human pose in RGB image or video.

Keywords **inverse kinematics, human posing, autoencoder neural network**

MSC(2010) **34A34, 92B20, 53A17**

doi: 10.1360/SSM-2019-0335