

NeuralTailor: Reconstructing Sewing Pattern Structures from 3D Point Clouds of Garments

Korosteleva & Lee

输入：三维点云；输出：二维服装缝制样式（sewing pattern）

文章Overview

如图1，从服装的点云中提取出缝制样式和拼接信息。

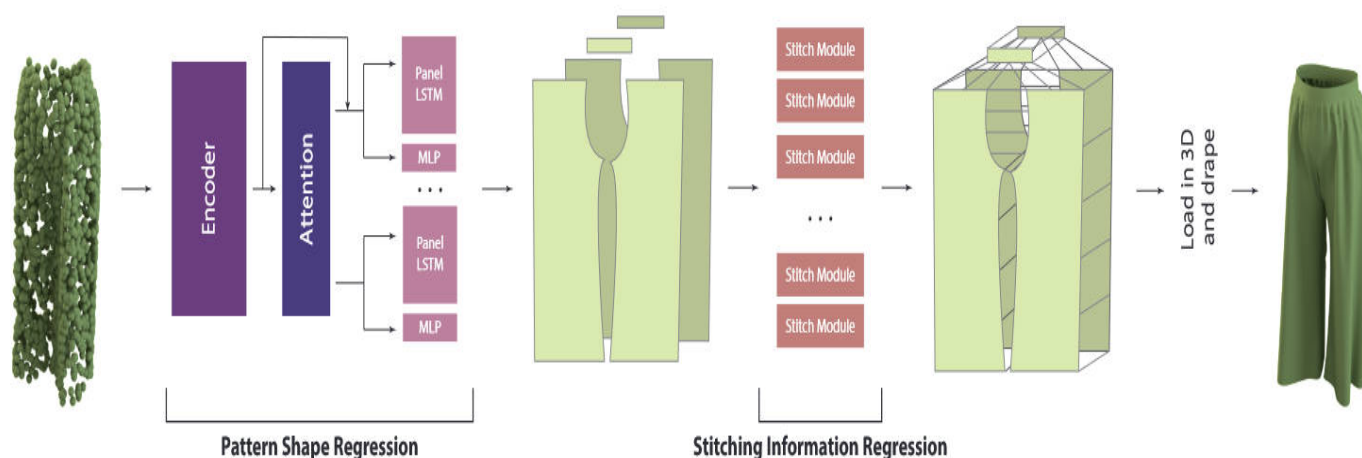


图1：文章框架

概念介绍

缝制样式

文章提出了使用**缝制样式**（**sewing pattern**）来作为服装静置状态（rest pose）时的基础表征。不同于以往基于参数化模板的方式，缝制样式能统一化描述各种服装类型和设计，是一个基于学习可变形物体结构恢复的案例，而不是把服装当成一个刚体对象进行研究。

一个缝制样式是一堆 2D 织物（文中将之称为**片 panel**）的集合，包括每个片在人体周围的**位置**（**placement**）以及如何把片拼接成最终服装的**拼接信息**。一个片会被建模成一个封闭的分段曲线，其每个分段（即每条边）是一条直线或一条 Bézier 样条曲线。缝制样式与现实世界服装的构造方式非常

接近，因此可以作为强大的**先验**信息来解决**数据问题**，比如静置姿态服装的形状必须根据物理形变得到，以及数据采集过程不完善的问题。

1. 数据集

本文使用的是按 [Generating Datasets of 3D Garments with Sewing Patterns](#) 生成的 **带缝制样式的 3D 服装数据集**。

该数据集涵盖了各种服装设计，包括T恤、夹克、裤子、裙子、连衣裤和连衣裙的各种变体，共计从19种基本类别中抽样了22,000件服装。每个服装样本包含以下内容：

- 一个适配 SMPL 中 T 型平均女性人体形状的 3D 服装模型。
- 服装对应的缝制样式。
- 一个模仿了一些扫描伪影（artifact）的有损 3D 模型。

该数据集在表示人体姿态和形状上能力很有限，但提供了大范围的服装设计内容。

片的分类

原始数据集会存在不同服装类型有相似片情况的存在，如裤子和连体裤都有 pant panel。因此根据服装类型的角色和位置对片类别进行了分类，如盖住T恤、连衣裙、连衣裤等服装的前面部分的片会被分到**前片（front panel）**类中。

附加样本滤波

原始数据集还会存在包含重叠设计（overlapping design）的服装样本。如图2，这些样本具有不同的缝制样式拓扑，可能属于不同的服装类型，但在悬垂（drape）时产生相似的 3D 形状。这种情况在现实世界中是很常见但很棘手的问题。为了把注意力放在缝制样式重建和拓扑泛化上，所以作者手动分析了数据集基础模板的参数空间，并过滤他们以得到基本上没有重叠设计的样本。

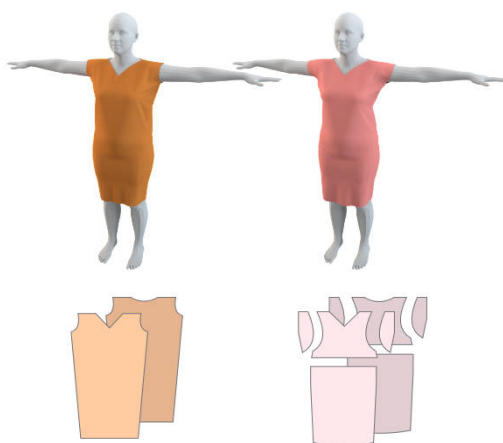


图2：重叠设计

数据集分割

带缝制样式的 3D 服装数据集根据服装类型分为训练集和测试集。

- 测试集中有7种类型在训练时仅作为评价使用。
- 每一组类的训练集中的100个样本作为模型选择的验证集，
- 每一组类的训练集中还有100个样本作为测试集，来比较模型在未参与训练服装类型数据上的性能。

这种分割方法在完整训练集中留下了19,236个服装样本，在采样滤波时的训练样本个数为 9,768。

2. 模型介绍

作者首先精心设计了一个原始基准模型来表征和重建缝制样式结构。该模型从输入点云中提取隐空间向量，然后通过 RNN 的两步结构将隐向量解码为缝制样式和表征为单个片所有边属性的拼接信息。

随后该基准模型被更新为NeuralTailor。在基准模型的基础上，作者引入了一个点级（point-level）注意力机制，来基于局部而非全局内容来评价单个片的隐码。然后将拼接预测拆分成一个独立的模块，来对边对（edge pair）分类成连接或不连接类别。

如第5节所述，这些改进使整个框架能泛化到训练时未使用的缝制样式拓扑上。实现的泛化特征对类似服装的多样化领域至关重要，因为手机一个完全典型的服装数据集是非常困难的，而通过训练特化单类别模型数据很难达成泛化设计。

3. 基准模型

整个基准模型的框架如图3所示：

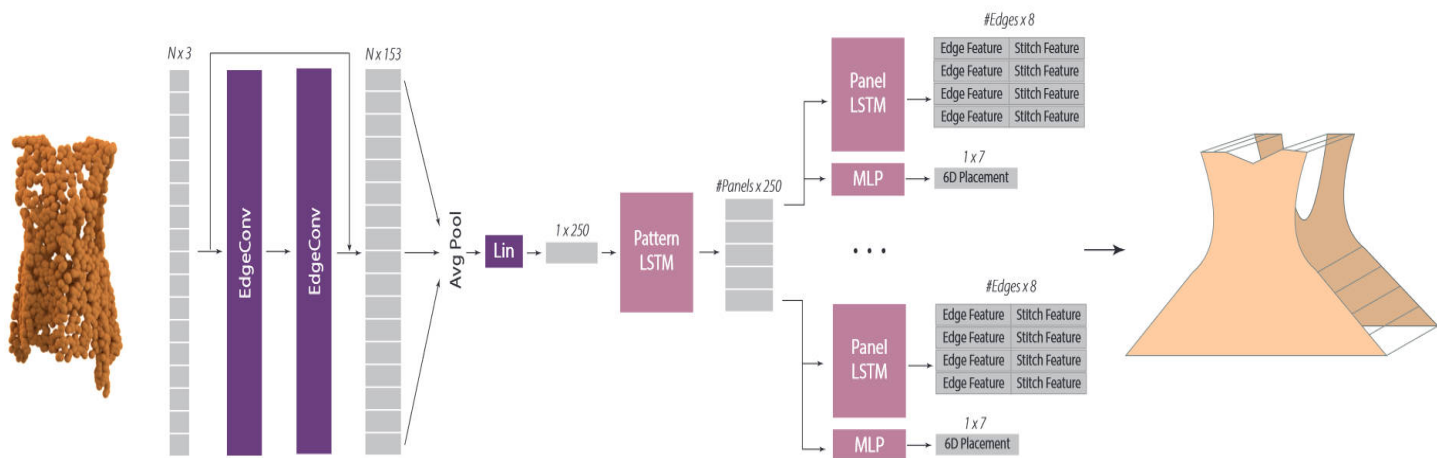


图3：基准模型框架

3.1 点特征编码器

作者使用EdgeConv作为编码器的基础模块，因为其简单并且性能和 SOTA 相当。EdgeConv 的主要优势是在特征空间中聚合信息，而不是通过在每个 EdgeConv 层上动态地重建连接图。如上框架图左边部分所示，文中的编码器由两个 EdgeConv 层和一个从 3D 点云输入到末层 EdgeConv 输出的残差连接组成，再通过一个平均池化层将每个点的特征聚合成一个单特征向量。

3.2 片编码的长短期记忆网络（LSTM）模块

完成信息聚合之后，该模型会根据全局隐码对每个片重建缝制样式的掩码。片之间没有顺序地组成一个集合，其基数（片数量）在不同的服装上是不同的。作者按照OrderlessRNN的思路，将片编码预测过程设计成一个 LSTM，因为它能建模对可变基数集。该模块**输入**服装全局隐码，**输出**片的隐向量序列，然后由片解码器处理成片形状。

3.3 片解码器

片解码器会从 LSTM 模块输出的片隐向量序列（包括组成片的边特征和一个回归片 3D 位置的附加线性模块）恢复出片的形状和拼接信息。

3.3.1 片的表征

片被建模为光滑封闭分段曲线，由一组边系列构成，这些边是直线或二次 B 样条曲线。使用样条曲线而非离散化表达是为了是表征更为简单，可以防止出现由分辨率导致的伪影，且可以确保拼接是单边对单边的连接。

3.3.2 边特征

为了表征片边的顺序，片解码器会把每条边输出为一个从起点 i 到终点 j 的 2D 边向量：

$$\vec{e}_{i,j} = v_j - v_i$$

其中 v_i 和 v_j 是点 i 和点 j 的 2D 局部坐标 (local coordinates) 。

如片的表征中所言，每个片都是一个封闭的分段曲线，其边在排序时形成一个环，片局部空的原点被定为环的第一个点。这样一来，任意点的 2D 坐标都可以通过片上前一个点的坐标与对应的边向量相加得到。

因为数据集在设计时保证了边环第一个顶点的选择和片之间边环遍历方向的一致性，所以在评价损失的过程中，只需要简单地使用数据中给出的边环顺序即可。

由于边不一定是直线，所以使用了**曲率坐标 (curvature coordinates)** 作为附加的边向量特征。曲率坐标是定义在边局部空间中的二次 B-spline 曲线控制点 (c_x, c_y) 。在这个局部坐标系内， $(0, 0)$ 和 $(1, 0)$ 指代边的顶点位置。因此， c_x 指代沿着边的位置，大致对应曲率峰 (curvature peak) 的位置， c_y 控制曲率的深度，直线边的曲率坐标则会被标记为 $(0, 0)$ 。

这样边特征可以被综合表示为

$$(e_x, e_y, c_x, c_y)$$

其中 $(e_x, e_y) = \vec{e}_{i,j}$ 是 2D 边向量坐标， (c_x, c_y) 是曲率坐标。

由于不同的片具有不同数量的边，每个片的边序列会进行零特征向量填充到大于或等于训练集中的最大边数，在本实验中最大值为 14。

3.3.3 拼接预测使用的拼接标签 (stitch tag)

拼接是网络输出中边之间的交叉连接。作者最初想把拼接信息直接包含到边特征中，将每条边的拼接信息定义为特征向量：

$$(f_{0/1}, s_1, s_2, s_3)$$

其中 $f_{0/1}$ 是一个**二元类 (binary class)**，代表该边是自由边还是需要拼接；拼接 tag (s_1, s_2, s_3) 是一个学习向量，指示需要完成拼接的边。

来自同一拼接的边标签应该是相似的，但不同拼接的边应该保持有差额的 tag。作者利用 tag 之间的欧氏距离来作为相似度量。**连接性重构 (connectivity reconstruction) 需要过滤掉自由边和连接的边，并比较连接边对 (pairs of connected edges) 的拼接 tag。**需要注意的是“自由”边不需要有意义的拼接 tag。

这样做可以得到一个简单的样式连通性表征，这种表征不依赖于拼接的数量或样式内边的数量，并且避免了显式引用边的 ID，允许编码多种不同缝制样式拓扑。网络就可以通过**损失函数**来学习提供能在训练

过程中强制修正表现的拼接 tag 的值。

3.3.4 3D 片位置表征

片在世界坐标系中的位置 (placement) 可以被表示为：

$$(q_1, q_2, q_3, q_4, t_1, t_2, t_3)$$

其中 (q_1, q_2, q_3, q_4) 是片的旋转四元数；片的位移 (t_1, t_2, t_3) 被表征为片的 2D 边界框顶部中点的 3D 位移。在大部分情况下，这个点对应于最影响片位置的身体特征（如：颈部、腰部），因此在特定的风格选择（如：裙子长度）中表现出稳定性。经过测试，相比于直接使用片本地空间原点作为参考点，这样操作能实现更好的 3D 位置预测。

3.4 损失函数

训练片形状和位置预测模块的完整损失为：

$$L_{total} = L_{edge} + L_{loop} + L_{placement} + L_{stitches}$$

- 边损失 L_{edge} 被用来评估片集合形状预测的质量。真实 (Ground Truth) 片表征首先会被[转换](#)成顺序 2D 边向量的形式。然后 L_{edge} 即等于 GT 数据和对应 NeuralTailor 输出的边特征之间，边向量和曲率坐标的均方误差 (MSE)。
- 环损失 L_{loop} 是为了保证片的封闭属性。它等于片的边序列起点和终点之间距离的 L_2 范数。
- 位置损失 $L_{placement}$ 是从 GT 位置信息转换为匹配网络输出规格的旋转和位移的 MSE。
- 拼接预测损失 $L_{stitches} = L_{class} + L_{tags}$ 。
 - L_{class} 被建模为一个二元交叉熵损失，将边分为自由类和非自由类。
 - $L_{tags} = L_{similarity} + L_{separation}$ 通过引用数据中的拼接信息来使输出的拼接符合[定义](#)，由两部分组成：相似度 (similarity) 损失和分离 (separation) 损失。
 - $L_{similarity}$ 鼓励一对拼接边尽可能的接近彼此。

$$L_{similarity} = \sum_{(i,j) \in stitches} \|tag_i - tag_j\|^2$$

- $L_{separation}$ 将来自不同拼接的拼接 tag 按预定义的间距 δ 分开：

$$L_{separation} = \sum_{\substack{i,j \in non_free \\ (i,j) \notin stitches}} \max(\delta - \|tag_i - tag_j\|^2, 0)$$

其中， i, j 是边的ID， $stitches$ 指所有要拼接的边对的集合， non_free 指参与任意拼接的非自由边的集合。这两个集合都是从 GT 缝制样式中得到的。

作者通过实验发现，在训练时，允许模型先学习到缝制样式的整体概念，经过几个 epoch 之后再引入 L_{class} 和 L_{tags} 能更高效，在作者的实验中，这几个损失是在40个 epoch 之后才引入的。

片排序和片填充

对上述损失的评价需要在缝制样式中选择片顺序。为了确保网络输出的片序列和同个类的 GT 片位置匹配，片需要以固定顺序排列成片向量，再对缝制样式中没有的类进行零填充，以补齐到与实际片相同的维度。这种填充策略允许不同拓扑的排序更加一致，并使网络更容易学习到同一类片的相似性。片之间的距离是由片中所有边特征和 6D 位置向量的桥接结果之间的欧氏距离表征的。

4. NeuralTailor: 改进泛化能力

在[基准模型](#)的基础上，作者进行了两项修改来鼓励结构上的模块化和基于本地内容（local context）的推理（reasoning）。

其总体框架如图4所示。相较于基准模型，NeuralTailor 引入了一个注意力 MLP 模块来预测每个点所属片类的注意力分数（MLP 权重是所有点共享的），然后在点特征聚合成每个片的掩码之前，这个分数会被用来衡量点特征。此外，NeuralTailor 使用一个分离的 StitchMLP 模块来还原拼接信息，这个模块将预测样式的边对分类成“被拼接”或“不被拼接”（MLP权重是所有边对共享的）。

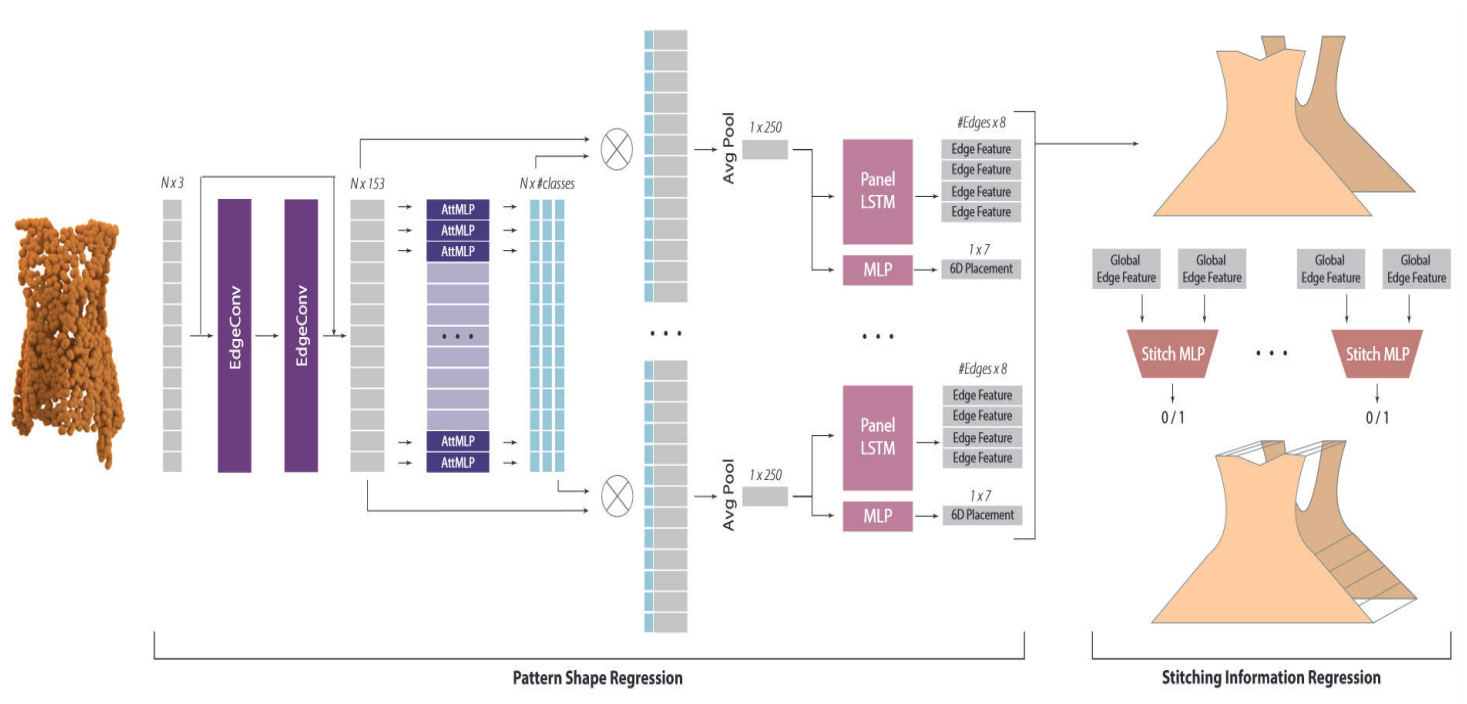


图4: NeuralTailor 框架

4.1 基于注意力的片编码

基准模型利用来自全局服装隐码的序列预测来重建每个片的隐码。这种全局制约让模型更倾向于依赖服装的整体形状而非每个部件的结构。

另一种选择是只关注输入点云的相关部分来重建对应的片的隐码。这种方法允许模型从相关的碎片中重构最终的缝制样式。另需考虑的是不同类型的服装由不同数量和类型的片。

根据**片的分类**，一个缝制样式中，属于同个片类的不同片不会同时出现。因此如上图，NeuralTailor 引入了一个额外的 MLP 模块来实现注意力机制，该模块作用于点特征向量并且预测了每个片类上每个点的概率分数，代表一个特定点属于给定片类的可能性。然后该模型通过相关类的注意力分数加权的**点特征编码器**，来简单的池化点特征。

理想情况下的注意力权重只会鼓励识别不同服装之间的片类对应的组件。所以注意力权重应该只用来突出每个片类相关的最小本地内容（context）。这样一来权重是稀疏的，并且每个点应该只参与少量的类。NeuralTailor 采用了 **SparseMax** 作为评价每个点注意力分数的注意力模块的最后一层。

4.2 拼接信息回归

基准模型基于拼接 tag 的模式拼接预测产生了拼接的有效表征，来重建单模型的完整缝制样式。然而其拼接信息和片形状都是从相同的片隐码从推断而来，导致两者存在联系，而训练数据里存在的拼接信息不完整会导致网络学习到训练数据的这一误差，产生错误的预测。出于这个原因，作者将片边不再作为片的部分，而作为独立的对象独立处理。实验表明，不需要输入几何形状信息，样式几何（pattern geometry）和片位置（panel placement）可以提供足够的信息来预测拼接。

如图4，作者构建了一个简单的 MLP 模型，以一组缝制样式边作为输入，输出这些边是否拼接的可能性。每条边都被表征为一个向量：

$$(v_x^{start}, v_y^{start}, v_z^{start}, v_x^{end}, v_y^{end}, v_z^{end}, c_x, c_y)$$

其中 $(v_x^{start}, v_y^{start}, v_z^{start})$ 和 $(v_x^{end}, v_y^{end}, v_z^{end})$ 人体模型片连接边顶点的坐标， (c_x, c_y) 是边的曲率控制点坐标，在**边特征**中介绍。

训练集结构

朴素（naive）训练集包括了服装数据集中每个缝制样式所有可能的边对组合。但这个训练集中大部分边对并不需要拼接。并且其缺少含大量片的复杂样式，因为边对的数量与缝制样式内边的总数成二次增长。这导致在添加更多的缝制样式进训练集时，训练集大小会快速增长。

为了解决这一问题，作者在每个 epoch 中从每个缝制样式采样固定数量的边对，对拼接的边对进行过采样，对不连接的边对进行欠采样。由于数据集包含很多共享缝制样式拓扑的样本，作者希望网络在训练时能为非连接边对获取足够的线索（clue）。此外还通过在训练时随机化这些属性来避免特定的顶点或边顺序造成的偏差。

4.3 训练过程的调整

NeuralTailor 框架的训练分为两步：

1. 训练模型进行缝制形状的回归。
2. 按顺序训练拼接预测模型。拼接预测模型是在样式形状 (pattern shape) 模型重建的边特征上训练的，而非用 GT 缝制样式的边。这样增加了推理时对样式形状模型输出中噪声的鲁棒性。

因为拼接预测模块被独立出来，所以模型的总损失为：

$$L_{patternshape} = L_{edge} + L_{loop} + L_{placement}$$

其中 L_{edge} , L_{loop} 和 $L_{placement}$ 和[基准模型损失函数](#)中是一样的。用作损失评价的 GT 片使用和基准模型训练时相同的方法。而新加的拼接回归模型被处理为使用二元交叉熵损失 (BCE) 的二叉分类任务。

5. 评价

5.1 评价指标

- 每个预测样式 (**#Panels**) 中片数量和每个预测片 (**#Edges**) 中边数量的准确率。
- 不回到起点的片环会被认定为边数不正确，这种情况下通常需要添加一条边来产生一个连接的形状。
- 用预测片和 GT 片的顶点之间的欧氏距离的 L2 范数来评估片形状预测的质量，曲率坐标会转换成片空间的片点 (**Panel L2**) 。
- 类似的方法计算片旋转 (**Rot L2**) 和位移 (**Transl L2**) 和 GT 差值的 L2 范数。
- 用预测拼接的平均精度 (**Precision**) 和召回率 (**Recall**) 来衡量拼接信息的质量

5.2 比较 LSTM 和基于注意力的样式形状复原方案

实验比较了基于全局服装隐码的基准分层 LSTM (**LSTM**) 架构和基于注意力模型的样式形状复原 (**Att**) 。

输出中的拼接 tag 会同时影响两个模型