数据集整理

模型训练使用的数据

- 格式约定
  - 时序数据
  - 图像坐标
  - 相机坐标系
- 数据约定
  - 图像地址imgs_path
  - 翻转指示位flip
  - 图像块patches
  - 图像块裁剪框patch_bbox
  - 手部检测框hand_bbox
  - 手部关键点二维位置joint_img
  - 手部关键点在图像块中的位置joint_patch_bbox
  - 手部关键点在手部检测框中的位置joint_hand_bbox
  - 手部关键点在相机坐标系中的位置joint_cam
  - 手部关节点的相对位置标注joint_rel
  - 手部关键点的有效性标注joint_valid
  - MANO姿态参数标注mano_pose

模型训练使用的数据

格式约定

时序数据

每一个应当包含对同一只手的连续 `T` 帧的原始图像信息以及标注信息。

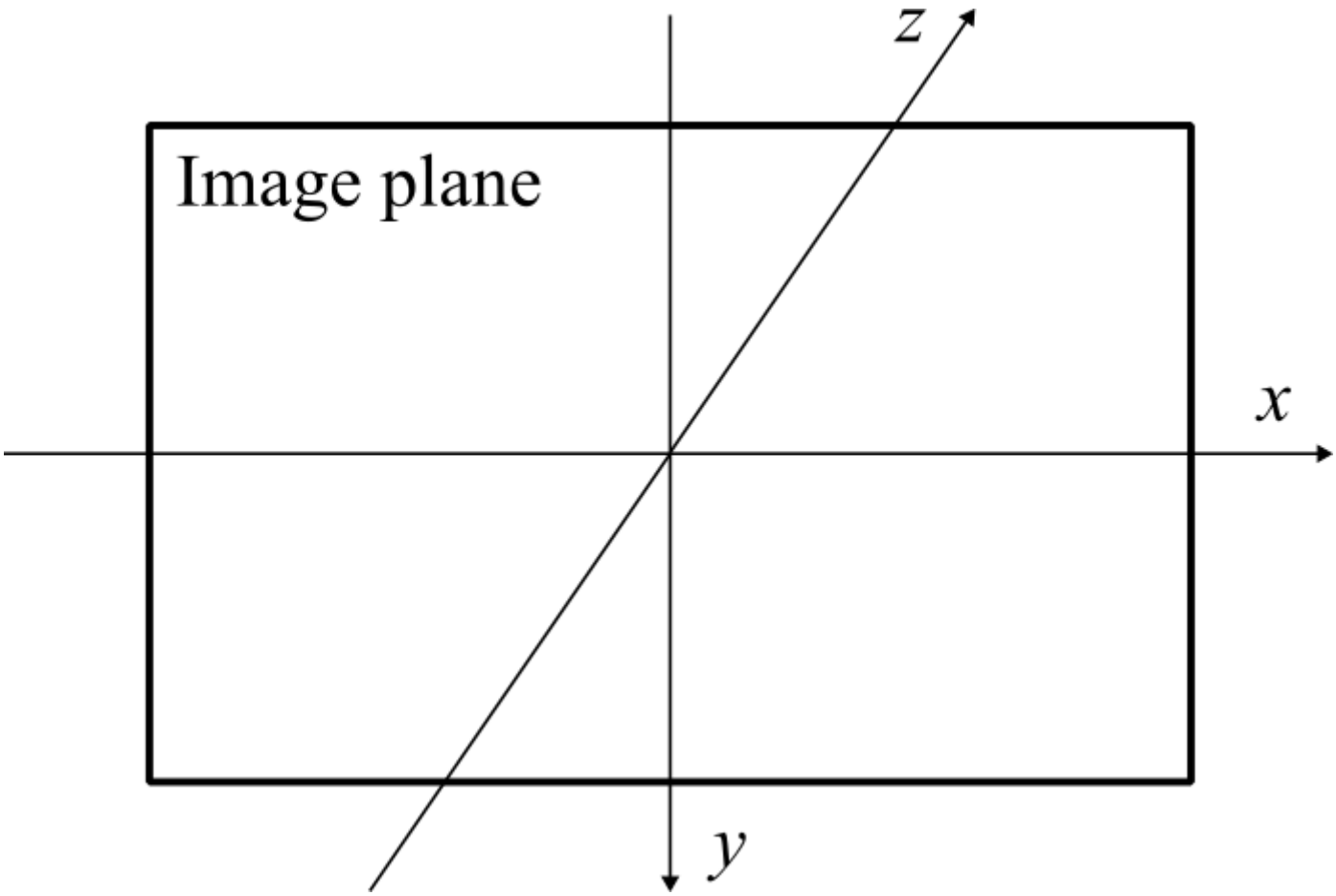
图像坐标

图像坐标 `(x,y)` 的约定为

1. 图像的左上角表示坐标 `(0,0)`

2. `y` 自上而下增大，范围为 `0~H-1`； `x` 自左向右增大，范围为 `0~W-1`

相机坐标系



数据约定

图像地址 `imgs_path`

格式为 `List[str]`，是长度为 `T` 的字符串序列，每个字符串为图像文件的地址，构成一段长度为 `T` 的图像序列。

翻转指示位 `flip`

格式为单一的 `bool`，用于表示这个序列中的标注的手是左手还是右手，如果是左手则为 `True`，否则位 `False`。因为我们会把左手图像反转为右手便于提升训练的数据量。

图像块 `patches`

格式为 `Tensor[T,C,H,W]`，这个是真正输入到网络中的图像张量。一般是从原始图像中的手部检测框中数据增强+截取+归一化+预处理得到的，其中 `H,W` 为其输入到模型的图像大小，常见的取值为 `224` 和 `256`。

图像块裁剪框 `patch_bbox`

格式为 `Tensor[T,4]`，是指在对图像进行了数据增强之后，对手部图像进行截取的框定位置，它是从原始的手部检测框（一般是紧的手部检测）进行正方形化+扩增得到的。通过这个检测框的位置可以

获得准确的透视信息嵌入。

手部检测框 `hand_bbox`

格式为 `Tensor[T,4(xyxy)]`，这个是指对图像进行了数据增强变换之后，手部裁剪框在新的图像上的位置，其长宽比是不规范的，用于框定手部检测的最小位置。**这里的 `x` 表示水平方向上的位移，`y` 表示垂直方向上的位移。**

手部关键点二维位置 `joint_img`

格式为 `Tensor[T,J,2]`，描述了手部的关键点在图像上的二维位置。其中坐标为图像坐标，约定见上。

手部关键点在图像块中的位置 `joint_patch_bbox`

格式为 `Tensor[T,J,2]`，描述了手部关键点在 `patches` 图像中的二维位置。其中坐标为图像坐标。

手部关键点在手部检测框中的位置 `joint_hand_bbox`

格式为 `Tensor[T,J,2]`，描述了手部关键点在 `hand_bbox` 能够裁剪出的图像中的二维位置。其中坐标为图像坐标。

手部关键点在相机坐标系中的位置 `joint_cam`

格式为 `Tensor[T,J,3]`，描述了手部关键点在相机空间中的位置，其中坐标为三维坐标，坐标系见上，单位为**毫米**。

手部关节点的相对位置标注 `joint_rel`

格式为 `Tensor[T,J,3]`，描述了手部关键点相对于根关节的位置，定义为

$$J_i^{\text{rel}} = J_i^{\text{cam}} - J_{\text{root}}^{\text{cam}}$$

手部关键点的有效性标注 `joint_valid`

格式为 `Tensor[T,J]`，描述了 `joint_img`、`joint_patch_bbox`、`joint_cam`、`joint_rel` 标注的每个关节点的有效性。如有效则为 `1` 否则为 `0`。

MANO姿态参数标注 `mano_pose`

格式为 `Tensor[T,48]`，描述了MANO手部模型的所有关节点的旋转。其中第 `3i~3(i+1)` 描述了第 `i` 个关节点的轴角旋转（默认MANO的坐标系）。注意该参数没有进行PCA处理，也没有提前减去姿态参数向量的均值。

MANO形态参数标注 `mano_shape`

格式为 `Tensor[T,10]`，描述了MANO手部模型的所有关节点的形态。该参数不应进行任何预处理。

MANO有效性标注 `mano_valid`

格式为单一的 `bool`，表示该序列的MANO标注是否有效。

序列标注时间戳 `timestamp`

格式为 `Tensor[T]`，描述了每一个标注的捕获时间戳，单位为**毫秒**。

焦距标注 `focal` 与主点标注 `princpt`

格式均为 `Tensor[T,2]`，描述了每一帧捕获的时候使用的相机内部参数，其中焦距为 `(fx,fy)`，主点为 `(cx,cy)`，相机内参矩阵定义为

$$\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

数据来源 `dataset_name`

格式为 `str`，描述了这一个序列的数据来自于哪一个数据集。

数据增强与给模型之前的预处理（数据规整）

数据增强应该在数据规整之前完成，因为涉及到数据的一致性变化，但是两者也存在可能的交织情况，因为对图像的增强变换可能产生裁切，导致在生成patch的时候图像全黑的问题。

在此首先整理需要实现的数据增强：

1. 三维几何的变化增强
    a. 旋转变换：图像围绕主点的旋转变换——场景中物体围绕z轴的旋转变换。变换参数为 $\alpha \in [0, 2\pi]$，为绕z轴正方向右手指向下旋转的弧度，对应到图像中则为绕着主点顺时针旋转的弧度
    b. 缩放变换：图像围绕主点的缩放变换——场景中物体沿着z轴的平移变换。变换参数为 $s \in [a, b]$，为z轴分量乘以的系数，对应到图像中则为围绕主点进行缩放 $1/s$
2. 图像整体的变换增强
    a. 内参的变化增强。变换过程如下
        i. 随机一个焦距缩放系数 $t \in [x, y]$，该系数乘以两个焦距系数得到新的焦距系数 $f'_x = sf_x, f'_y = sf_y$
        ii. 在 $\mathcal{N}(\mathbf{c}, (\|\mathbf{c}\|/9)^2)$ 中随机采样一点作为新的主点系数
        iii. 新内参记作 K'
        iv. 使用透视变换 $K'K^{-1}$ 对图像以及二维标注进行变换
3. 像素层面的图像增强

针对三维标注要进行的数据增强变换

旋转变换：$\begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$

位移变换：$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & s \end{bmatrix}$

针对二维坐标和图像要进行的数据增强变换

旋转变换：$\begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$

缩放变换：$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/s \end{bmatrix}$

内参引起的图像变换：$K'K^{-1}$

预先转换好的数据

预处理好的数据最好先尽可能在保留原始数据的同时兼容模型的数据格式，设计如下：

1. 由于姿态数据集基本上都是按照clips进行存储的，所以预处理好的数据也按照clips进行组织，长度就是原始数据集中每个clips的长度
2. 使用webdataset的格式进行数据组织
3. 每一个sample包含上述的所有项，但是需要将 `torch.Tensor` 替换为 `numpy.ndarray`
4. 添加一项**原始图片** **`imgs`** ，格式为 `ndarray[T,C,H,W]` ，为原始图片
5. 添加一项**附加信息** **`additional_desc`** ，格式为 `List[str]` ，用于存储这个数据的部分原始信息
6. 存储手性 `handedness` 而不是 `flip`

图像地址 `imgs_path`

格式为 `List[str]` ，是长度为 `T` 的字符串序列，每个字符串为图像文件的地址，构成一段长度为 `T` 的图像序列。

原始图片 `imgs`

格式为 `List[Bytes]` ，是长度为 `T` 的图像编码序列，每个编码为原始图像使用WebG的编码结果。

手性 `handedness`

格式为 `str`，用于表示这个序列中的标注的手是左手 `l` 还是右手 `r`。

手部检测框 `hand_bbox`

格式为 `ndarray[T,4]`，这个是指对图像进行了数据增强变换之后，手部裁剪框在新的图像上的位置，其长宽比是不规范的，用于框定手部检测的最小位置。

手部关键点二维位置 `joint_img`

格式为 `ndarray[T,J,2]`，描述了手部的关键点在图像上的二维位置。其中坐标为图像坐标，约定见上。

手部关键点在手部检测框中的位置 `joint_hand_bbox`

格式为 `ndarray[T,J,2]`，描述了手部关键点在 `hand_bbox` 能够裁剪出的图像中的二维位置。其中坐标为图像坐标。

手部关键点在相机坐标系中的位置 `joint_cam`

格式为 `ndarray[T,J,3]`，描述了手部关键点在相机空间中的位置，其中坐标为三维坐标，坐标系见上，单位为**毫米**。

手部关节点的相对位置标注 `joint_rel`

格式为 `ndarray[T,J,3]`，描述了手部关键点相对于根关节的位置，定义为

$$J_i^{\text{rel}} = J_i^{\text{cam}} - J_{\text{root}}^{\text{cam}}$$

手部关键点的有效性标注 `joint_valid`

格式为 `ndarray[T,J]`，描述了 `joint_img`、`joint_patch_bbox`、`joint_cam`、`joint_rel` 标注的每个关节点的有效性。如有效则为 `1` 否则为 `0`。

MANO姿态参数标注 `mano_pose`

格式为 `ndarray[T,48]`，描述了MANO手部模型的所有关节点的旋转。其中第 `3i~3(i+1)` 描述了第 `i` 个关节点的轴角旋转（默认MANO的坐标系）。注意该参数没有进行PCA处理，也没有提前减去姿态参数向量的均值。

MANO形态参数标注 `mano_shape`

格式为 `ndarray[T,10]`，描述了MANO手部模型的所有关节点的形态。该参数不应进行任何预处理。

MANO有效性标注 `mano_valid`

格式为单一的 `bool`，表示该序列的MANO标注是否有效。

序列标注时间戳 `timestamp`

格式为 `ndarray[T]`，描述了每一个标注的捕获时间戳，单位为**毫秒**。

焦距标注 `focal` 与主点标注 `princpt`

格式均为 `ndarray[T,2]`，描述了每一帧捕获的时候使用的相机内部参数，其中焦距为 `(fx,fy)`，主点为 `(cx,cy)`，相机内参矩阵定义为

$$\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

附加信息 `additional_desc`

格式为 `str`，用于为这个clips添加来自数据集的补充标注。