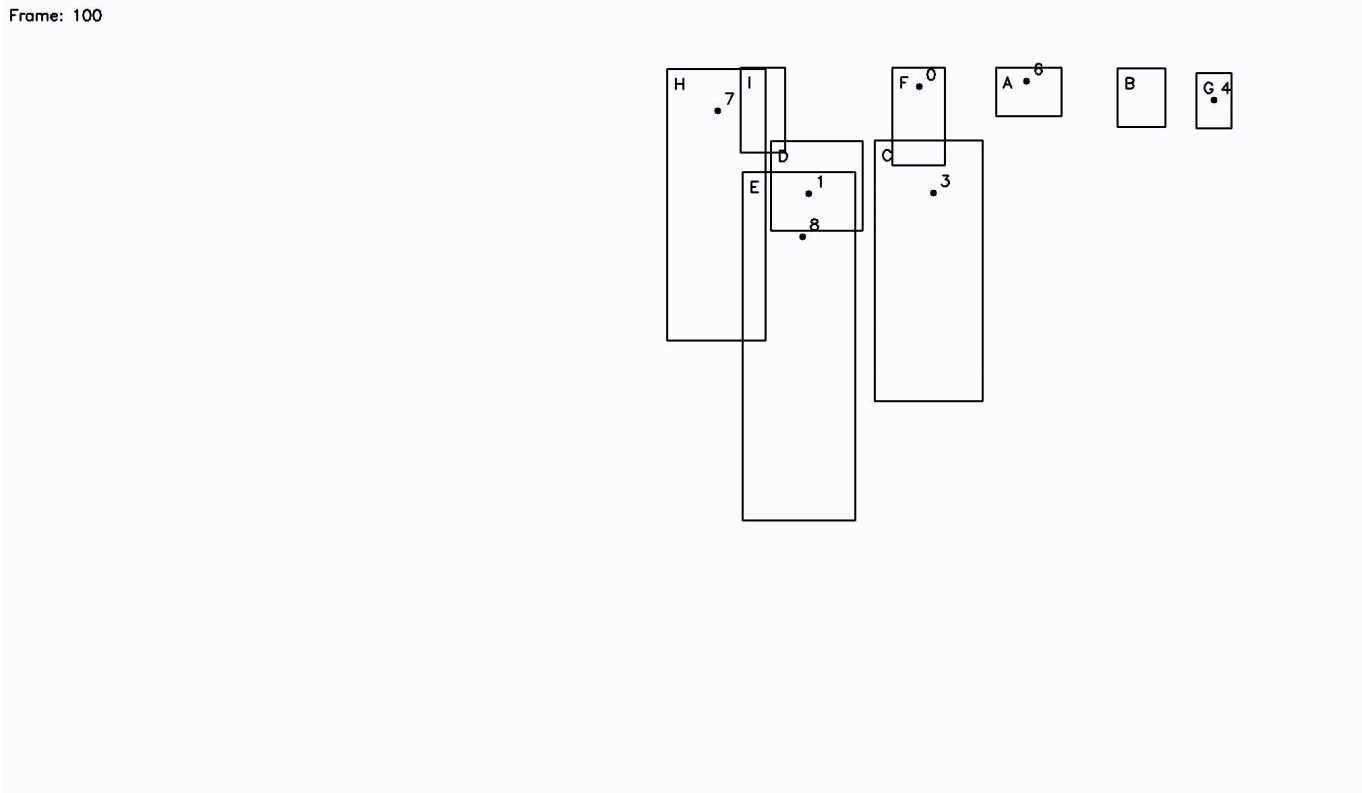
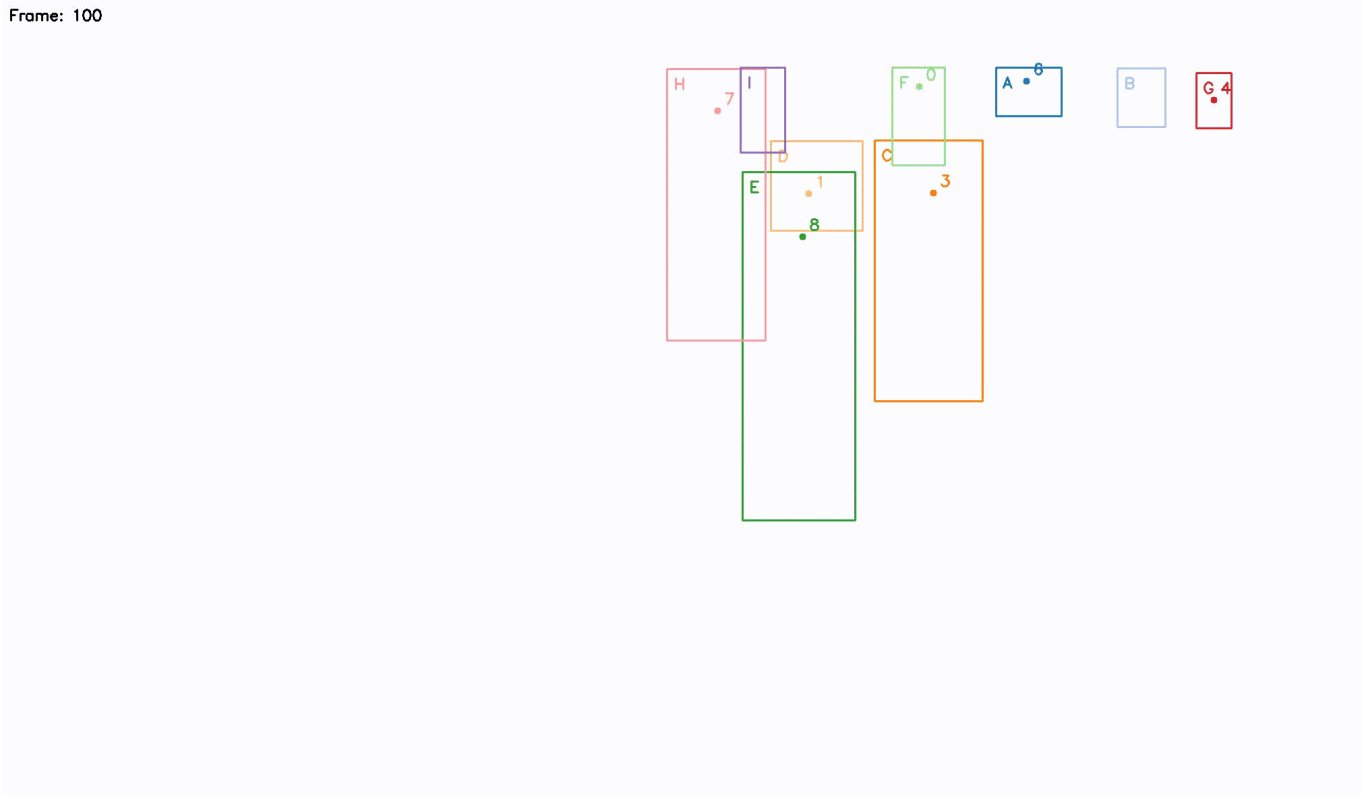
Labeling human position in videos

Two different tools are used for annotating human positions in videos: one for labeling bounding boxes and one for labeling neck joints. Each tool provides annotations (bounding box/neck joint) alongside person identity information for each video frame.

Since input data is provided by two different annotation tools we don't know if a neck joint and a bounding box belong to the same person.



The goal is to merge annotations from these tools and have consistent (global) person ID in both sets of annotations, i.e. the goal is to find which bounding boxes and neck joints belong to the same person.



Data sets

There are two data sets:

- *Public data set* is used for developing your solution. After you submit your solution, you will be able to see how well your solution performs against this data set. *Public data set* is not used for calculating the final score. Public data set is available **here**.
- *Private data set* is used for testing your solution. The final score will be measured against this data set. *Private data set* and the final score will be available after the homework finishes. *Private data set* contains different data than the *public data set*, but the type of data (video length, number of labels, bounding box sizes...) is roughly the same.

Problems with labels you will encounter

1. People tracks can intersect.
2. People can be located at different distances from the camera. People further away from the camera appear smaller, so the labeled bounding boxes will be smaller.
3. Sometimes, neck joint is not visible (and not labeled), while person is partially visible and bounding box is labeled. And vice versa, it can happen that only neck joint is labeled.
4. To speed up the labeling, only a subset of frames can be labeled. E.g. every 6th frame can be labeled to speed up labeling 6 times. To get position/bounding box of a person in frame that is not labeled, we can find two closest frames that are labeled and do simple interpolation.
5. There can be errors in labeling (either missing labels or completely wrong ones).

Input

Inputs are given through standard input.

Path to the json file with labeled bounding boxes is given as the first input line. Path to the json file with labeled neck joints is given as the second input line.

Example of input:

```
D:\data\public\set\0\bboxes.json
D:\data\public\set\0\joints.json
```

Input paths can be both absolute and relative paths.

If in doubt, please refer to the first task to see how to read the input.

Format of labels

Example of bounding box labels:

```
{
    "frames": [                         # Labeled frames
        {
            "frame_index": 6,           # Position of the frame in the video
            "bounding_boxes": [         # List of labeled bounding boxes
                {
                    "identity": "A",    # Person identity
                    "bounding_box": {   # Bounding box coordinates in the frame
                        "h": 0.061759259259259264,
                        "w": 0.03569791666666665,
                        "x": 0.48813541666666665,
                        "y": 0.000675925925925926
                    }
                },
                ...
            ],
            ...
        },
        ...
    ]
}
```

Example of neck joint labels:

```
{
    "frames": [                         # Labeled frames
        {
            "frame_index": 0,           # Position of the frame in the video
            "joints": [                 # List of labeled neck joints
                {
                    "identity": "0",    # Person identity
                    "joint": {          # Neck joint coordinates in the frame
```

```
                        "x": 0.6658313,
                        "y": 0.01939916145
                    }
                },
                ...
            ],
            ...
        },
        ...
    ]
}
```

For more details, please refer to the provided public data set.

Json files with labeled data consist of list of labeled video `frames`. Each frame contains information about its index in the video (`frame_index`) and contains a list of labels (`bounding_boxes` or `joints`).

- Labeled bounding boxes in the frame are stored in `bounding_boxes` list. Each bounding box contains person identity information in `identity` field and bounding box coordinates in `bounding_box` field. Person `identity` information is consistent across all frames of a given video.
- Labeled neck joints in the frame are stored in `joints` list. Each neck joint contains person identity information in `identity` field and neck coordinates in `joint` field. Person `identity` information is consistent across all frames of a given video.

All coordinates are normalized by corresponding video dimensions. `x`, `w` are normalized by video width; `y`, `h` are normalized by video height.

Notes:

- All video frames have the same dimensions.
- Person `identity` information is a string (an array of alpha numeric characters).
- Feel free to use `json` library for loading labeled json files.
- Feel free to use `pydantic` library for parsing of json files.

Output

Resulting list of matched person identities should be printed to the standard output, one pair per line. Single output line should consist of a column-separated pair (`joint_identity`, `bounding_box_identity`). Example of the output:

```
0:F
1:D
2:I
3:C
4:G
6:A
8:E
7:H
10:B
```

Notes:

- Results don't have to be sorted.
- Only matched identities should be printed. If you were unable to match identity (e.g. bounding box exists, but neck joint is not labeled), do not print it. Otherwise, test case will be scored with 0 points.

If in doubt, please refer to the first task to see how to print results to the output.

Available packages

- `numpy`
- `scipy`
- `pydantic`
- all packages from the standard python library

Scoring

- Correct result brings 25 points per test case.
- Result with only one error brings 15 points per test case.
- Other results don't bring any points.

Constraints

- Time limit is 2s.
- Memory limit is 80 MB.

If in doubt, please refer to the data from *public data set* and proceed with a reasonable assumption.