# 1 Approach

The overall goal is to identify the trajectories of microfluidic droplets, provided microscopic images across multiple time frames. On a high level, this comprises two main tasks. The first is to encode available information about each droplet in meaningful and efficient droplet representations. The second is to leverage these representations to perform the tracking across frames. Here, the main challenges are the large number of droplets per frame (up to hundreds of thousands of droplets), potentially large movement in between frames due to the low temporal resolution of the microscopy data, and the lack of groundtruth trackings for experiments with significant movement. The combination of these also differentiates this problem from many other multiple object tracking problems (paper references). Our proposed approach relies on a combination of classical object detection and convolutional neural networks (CNNs) for finding droplet representations and solving optimal transport (OT) problems for tracking. In the following, we outline the main ideas of the method and how to evaluate its performance. For theoretical motivations and implementation details, we refer the reader to the appendix. The code is available via...

## 1.1 Droplet Representations

A simple way of representing a droplet is to describe it using some n-dimensional real vector. The only requirements we impose on this representation are that it captures information that is useful to the tracking algorithm and that it is low-dimensional to ensure computational feasibility. Since solving the OT for tracking, to some extent, relies on matching droplets by the similarity in their representations between two consecutive frames, "useful" information may be properties of a droplet, which remain similar across frames. Hence, we want to avoid encoding features of droplets which change over time as they make identifying correct matches across frames more difficult. This also motivates our decision to discard the RGB channels for our method, as they may change over time by experimental design. Next, we describe how to extract desirable representations from the remaining brightfield and DAPI channels of the microscopic image data.

**Positional Information** For extracting the position of droplets within the image data, using an appropriately preprocessed brightfield channel proves sufficient. Since all droplets are almost perfectly circular, the Houghes Circles algorithm (ref!) is effective in finding the centers and radii of each droplet. The output is then refined using the RANSAC algorithm (ref!). This step provides the position for each droplet. For experiments with very little to no movement, this 2-dimensional representation of the droplets is already enough for near-perfect tracking of droplets. However, when droplets move significantly in between frames, we need to provide additional information.

**Visual Information** After obtaining the droplet positions, we can cut out image patches of size 40x40 pixels, each containing one droplet. Using the brightfield and DAPI channel this provides us with two image patches describing a single droplet for one frame. In principle, we could vectorize the pixel values of both images and concatenate them to obtain a visual feature representation, which would result in a (40*40*2)-dimensional representation. However, there may be changing visual features which we do not want to capture and the high dimensionality would make tracking computationally very costly. Hence, we use a deep learning approach to find much smaller, more compressed 20-dimensional representation vectors of the relevant visual information. To do so, we employ a combination of convolutional neural nets (CNNs) and contrastive learning. CNNs are a popular tool for efficiently learning low-dimensional representations of images and can be used for many different tasks involving image data. Here, contrastive learning refers to an approach to training that ensures that the model learns representations that specifically capture visual features that, for the same droplet, remain similar between frames. For training the model, we use image data with very little to no movement.

**Biological Information** In addition to positional and visual information, it is possible to represent droplets through their biological characteristics. For instance, in our model, we introduce the option of using the number of cells that can be obtained by special image analysis algorithms. Other properties such as cell types, barcodes et cetera can be of interest if there exist methods to extract them manually. In principle, the representations learned by the contrastive learning model may also capture these types of information, but less explicitly and maybe less efficiently so.

Finally, for each droplet, we concatenate the representations learned into a single 23-dimensional vector (2 for position, 20 for visual features, 1 for number of cells) and use them as the support for the optimal transport.

### 1.2 Tracking via Optimal Transport

To explain how to apply the theory of optimal transport to find good droplet trajectories, we consider the following situation: There are two frames, a source frame (current), and a target frame (next). Each droplet in the target frame corresponds to some droplet in the source frame but is 30 minutes older. From the previous section, we also know how to obtain the 23-dimensional representations for each droplet. Now the goal is to find a matching between droplets in the source and target frame which reflects the true droplet identities and also provides a measure of uncertainty of the algorithm. One way of solving this problem is modeling it as an OT problem.

**Optimal Transport** In mathematical terms, OT addresses the question of how to best transport mass from a given source distribution to a target distribution. A typical example to demonstrate the idea involves finding the optimal way of transporting resources from mines to factories, where the mines and factories are spread across the country and the cost of transportation is proportional to the distance between facilities. The solution of the optimal transport corresponds to a transport plan that has minimal transportation cost while making sure that all factories obtain as many resources as they need and that all mines deliver only as much as they produce.

**Droplet Matching** With this analogy, mines correspond to droplets in the source frame, while factories are droplets from the target frame, each represented by their respective droplet representations. The transportation cost is now a measure of dissimilarity between the representation vectors. Since the droplet representations consist of three different types of information, also the cost between them has three terms. The relative weighting of these terms is determined by tunable model parameters. For experiments with very little movement, it is intuitive to penalize large movements for the optimal transport and hence to weigh the positional cost heavily. For large movements, however, the positional information is not enough and therefore, the other cost terms may be weighted more strongly. Finding appropriate weights and suitable parameters for the algorithm solving the optimal transport is therefore crucial. The resulting optimal transport describes how much "mass" of a droplet in the source frame goes to different droplets in the target frame. The target droplet to which most mass is assigned is considered the match proposed by the algorithm and the normalized amount of mass transported is used as a measure of uncertainty for that matching.

**Trajecory Creation** Finally, the OT can be solved for each frame transition, and the proposed matchings annealed to obtain predicted droplet trajectories. Such a trajectory contains the proposed position of a droplet across time with an uncertainty measure for each transition and all possible sub-trajectories. These results can also be further filtered based on configurable criteria and to exclude merging trajectories, where a target droplet was matched with two or more source droplets. Lastly, we provide movement statistics which inform about the amount of movement of the proposed trajectories.

### 1.3 Evaluation

After running this tracking approach on experimental data, we provide two tools to visualize the predicted trajectories. One depicts the movement of all droplets, and the other shows the image patches in all channels across all frames. These are helpful to qualitatively assess the results from the algorithm. However, they are not practicable for finding optimal choices for the parameters of the model that need to be configured. Hence, it is useful to have access to experimental data with movement where the true trajectories are known - a gold standard dataset. Unfortunately, manually curating such ground truth is intractable, as it would require experts to track a myriad of droplets by eye, which is highly time-consuming and prone to errors. We came up with two approaches to finding an alternative to the gold standard. The first relies on simulated positional data, the second on using real data with high temporal resolution, i.e. a much higher number of frames for the same period.

**Simulated Silver Standard** The idea for the silver standard by simulation is to combine droplet image patches from microscopy data with hardly any movement, where the true trajectories are static and therefore known, with simulated positional data to introduce artificial movement. The simulation is a simple rigid-body simulation, which aims to imitate different modes of movement observed in real data. See Figures ... and refer to ... for implementation details.

**Experimental Silver Standard** Another approach to creating something close to a ground truth is based on the assumption that our tracking algorithm works well on data with very little movement. We run our tracking approach on experimental data with very high imaging frequency. This ensures that in between any two frames, there is only little movement. We partially check the quality of the trajectories by investigating the visualizations. The algorithm can then be evaluated on the same dataset, using far fewer frames for trajectory prediction.

**Evaluation Metrics & Model Calibration** Having access to silver standard datasets allows us to quantitatively assess the performance of our tracking approach and thereby configure model parameters optimally. Meaningful metrics include precision@k, the area under the precision-recall curve (AUPRC), and the Brier score (see ... ref!). Furthermore, the evaluation engenders the calibration of the uncertainty measures provided by the algorithm, i.e. aligning the model's uncertainty measure with the actual outcomes (see ... for details ref!).