Tracking and segmentation, a tool for assessment of human engineered heart tissue

Danjel Keekstra, *dkeekstra@outlook.com, Master Embedded Systems*

***Abstract*— Engineered heart tissues (EHT) are a promising *in vitro* 3D model for studying heart disease and drug discovery. This study presents a method to analyze microscopic videos of the EHTs automatically. The analysis consists of tracking the movement of the EHT, which is relatively small and therefore challenging, and segmentation of the different parts of the EHT where the difficulty is in the different illuminations and noise. This study proposes template matching with the extension of a novel method that delivers smooth sub-pixel precision trajectories for tracking. For segmentation, deep learning models are used to segment the EHTs. This study presents a framework to evaluate the quality of the tracking and segmentation to assess the outcome of the proposed methods.**

Keywords*—Tracking, Segmentation, Human-engineered heart tissue, Template matching, Sub pixel interpolation

I. Introduction

Cardiovascular diseases (CVDs) account for 4.1 million deaths annually in Europe alone [1]. Globally, CVDs are the leading cause of death [2]. To treat these diseases, human heart (patho)physiology and drug discovery studies are now widely done using human pluripotent stem cells[3]. In these studies, scientists use 2d and 3d models to mimic the heart. The most promising model to mimic the heart *in vitro* 3D is engineered heart tissues (EHTs) [3]. Scientists create EHTs by using human pluripotent stem cells (hPSCs) to derive cardiomyocytes (hPSC-CM) for the creation of human-engineered heart tissues (hEHTs) [4].

The Applied Stem Cell Technologies (AST) and BIOS Lab-on-a-Chip group from the University of Twente developed a platform that allows the creation of stem cell-derived cardiac tissues. In this platform, an EHT can be grown between two silicon pillars, as is shown in the brightfield in Figure 1. This platform allows scientists to analyze the behavior of EHTs. This analysis consists of capturing the contraction properties. Additionally, the surface area of the tissue is relevant biological information. To gather the contraction properties, the contraction movement of the tissue needs to be collected. The contraction movement of the tissue can be found in the movement of the two pillars that can be seen in Figure 1. Tracking the position of the pillars over time with sensors would be challenging. A video from the brightfield microscope could be manually analyzed. This manual analysis would be challenging and time-consuming. A software algorithm that can automatically analyze the EHTs by brightfield videos would be of great use for the scientists working with the EHTs. Many EHTs need to be analyzed, so accurate and fast analysis is needed. The main problem in automatically analyzing the brightfield videos lies in the present low contrast, noise, and low pillar displacement. Segmentation of the tissue in a brightfield image could supply the additional surface area information. The challenge in segmenting the surface of the tissue in the brightfields are the present noise and low contrast and the different orientation and shapes in each video.

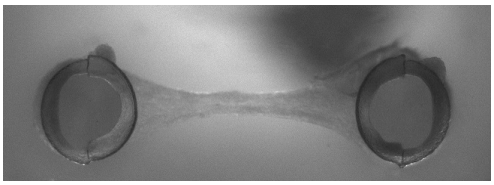

Fig. 1. Brightfield of a EHT with 2X magnification

For the analysis of the brightfield videos, this study presents tracking by template matching with sub-pixel precision to track the displacement of the pillars and semantic segmentation by deep learning to provide a template for the template matching algorithm and determination of the surface area of the EHT. In this study, we answer the following questions:

- (RQ1) Is template matching with sub-pixel precision a solution for tracking EHTs that produces correct results?
- (RQ2) Can deep learning-based segmentation methods be used to provide templates of the engineered heart tissues pillars for the template matching with the sub-pixel precision method?
- (RQ3) Can deep learning be used to determine the surface area of the EHT by segmenting the engineered heart tissue?
- (RQ4) How can we evaluate the performance of the tracking algorithms?
  - (RQ4.1) What are quantitative measures that can be used, and how would the methods score?
  - (RQ4.2) What qualitative characteristics of the tissues can be measured by analyzing the tracking algorithms output?

II. Related works

In the field of EHT analysis, not many publications are available. There are custom software tools, but it is not always clear how precise and efficient they are. In similar research fields, for instance, sperm cell research and Red Blood Cells in Urine research, multiple methods exist that may be promising for use on the EHTs videos. In the following subsections, we discuss these methods.

A. Tracking

In the field of cardiac research, multiple methods for tracking the movement of heart tissues are used. Some of the algorithms are based on a cell-level scope, and others detect the whole heart tissue movement. Researchers developed Musclemotion[5] to quantify cardiomyocytes and cardiac muscle contraction *in vivo and in vitro*. It uses subtraction of a reference frame to the current frame to determine the movement in a frame. The algorithm suffers from distortions, illumination problems, sample shifts, vibrations, and debris in the videos. These problems make the algorithm not robust in not ideal situations. The developers use filtering to solve these problems partly. However, this is not a good solution since filtering will lose information.

In Live-cell tracking [6], scale-invariant feature transform (SIFT) keypoint detection is used to track cells in microscopic images. As SIFT detects key points on the whole image, and an EHT moves differently within the tissue, using key points that are not on the pillars makes the average movement less accurate. This inaccuracy could be partly solved by segmenting the pillar first and only accepting key points within the pillar region. But it would still be compromised by the different movements within the tissue.

In motion vector analysis [7] and image-based evaluation of contraction–relaxation kinetics [8] the block matching algorithm (BMA) [9] is used to track the movement vectors of heart tissues. Their approach is to find the movement vectors of the EHTs by matching predefined-sized blocks from the current frame to the next frame. To gather the EHTs movement in a video, the region of interest of the EHT needs to be selected. The method then uses the corresponding movement vectors gathered from frame to frame to determine the movement of the whole region of interest. This method works well for analyzing parts of the tissue's movement. However, it suffers from areas where the movement is different, for instance, on the edges of the tissue. The block shape will take average movement and take the parts that do not belong to the tissue. Also, a good segmentation of the tissue is needed to do the tracking automatically. The movement in parts of the tissue can be different from the whole movement of the tissue. This difference in movement can also cause inaccuracies.

Another technique used in the field of tracking cardiac motion in EHTs is optical flow. In [10], [11] it is used to track the motion of heart muscle tissue. With optical flow, motion is tracked by following the pixel's intensity by differentiating in the horizontal and vertical direction and in time. The cardiac tissue has parts where the spontaneous contraction behaves out of the expected, affecting the total result of the movement. Although optical flow can give better insights into the tissue's behavior, it will struggle to find the total behavior of the tissue since the movement in parts of the tissue can be different from the whole movement of the tissue. Flow analysis of a whole video will also increase the computational time significantly.

Template matching is a known method to locate objects in the medical research field. In [12] template matching is used to locate organisms in microscopic images. Because the observed organisms are all slightly different multiple templates are used to get better performance. In [13] template matching is used to find the optical disk. Here a histogram of the colors in the image is used to create the correct template. The developers of [14] do segmentation to create a template for tracking by template matching. A reinforcement learning algorithm determines if the object's shape is changed too much to track the object. If the object changes too much, a new template is created to start tracking with template matching again. In [15] template matching is used to track the motion of sperm cells. Here the orientation of cells changes which makes it hard to track. Tracking with template matching suffers from the change of the object's orientation, illumination, and shape. In the proposed studies, researchers propose multiple solutions to overcome this. The EHT videos have the advantage that the pillar shape, illumination, and orientation stay similar. Therefore template matching is promising to use for tracking the movement of EHTs.

In [16] sperm cells are tracked by a modified version of the Discriminative Correlation Filter with Channel and Spatial Reliability (CSR-DCF) [17]. Here a filter is trained by deep learning to find the shape of the sperm cells in the microscopic pictures and then tracked. The method is precise in tracking sperm cells. A sperm cell moves on average 5 pixels per frame. In contrast, the movement of the EHTs is far below 1 pixel per frame. This makes that for this method, the tracking accuracy on the videos of the EHT could be much lower. The presented method is hard to set up since the model has to be trained with a ground truth data set, which is extensive to make.

In the field of tracking, there are many studies not focused on medical or microscopic images [18]. Although these studies use methods that may be suitable for tracking the EHT movement in the videos, we cannot compare with their results since the used material is different, and the results may differ.

B. Segmentation

To the best of our knowledge, there does not exist an algorithm specialized for segmenting EHTs. In related fields, researchers propose methods that may also be promising for segmenting the three regions of the EHTs.

In [19], [20] Hough circle detection[21] is used to segment red blood cells in microscope images. The method is also used in [22] to detect the location of the optical disk in the eye. Hough circle detection works on circle-shaped objects, but with other round-like shapes, it suffers. Since the pillars in the EHT videos are not perfectly round, and their shape may change, the Hough circle direction is not ideal.

State-of-the-art image segmentation methods are nowadays based on deep learning and convolutional networks. In [23], [24], [25] DL is used to segment objects in medical images. Since DL is an active and fast-growing research field, we can use many different techniques and models to segment the regions of the EHTs. In [26] a recent survey is done of the available DL models for segmentation. The results achieved in these studies show that although DL is not perfect, there lies a great potential for DL on the EHT segmentation task. U-net [27] is a deep learning model developed explicitly for efficiently segmenting biological microscopy images and

therefore promising to be used on brightfield frames of the EHT videos. Deeplabv3 with a resnet-101 backbone[28] scores well in the survey on the different data sets. Although it is not specially designed for biological microscopy images, it can be promising to try.

Another method used for the segmentation of medical images is thresh-holding and edge detection. These methods have been around for some time and are well known. In [29], [30] edge detection is used to segment red blood cells. Edge detection suffers from noise and is therefore not robust. The researchers in [31], [32], use thresholding to find blood cells by the difference of pixel intensity. The proposed methods use colored images, but gray-scale images like in the EHT videos should also work. Thresholding can suffer from differences in color between images, making it a less robust method.

There also exist many studies on segmentation not focused on medical or microscopic images. Although these studies use methods that may be suitable for the segmentation of the EHT videos, we cannot compare the results of these studies since the used material is different and the results may differ.

III. Materials

The EHTs used in this study are formed in the fluid around the two pillars. The pillars are silicon molded and are not symmetrical or round at the top, as can be seen in Figure 1. The shape of the pillar is cylindrical with on top a round-like plateau. The molding places the pillars at a distance of about 3.2 mm from each other. This distance is kept consistent across all the silicon molds.

The EHTs are recorded in an inverted microscope under temperature and humidity control using a high-speed camera with 100 frames per second (fps). The microscope outputs frames in an 11bit grayscale format. There are three relevant regions in the brightfield of an EHT, which can be seen in Figure 3. The red encircled part is the inner part of the silicon pillar. Because of the silicon molding, the shape is not perfectly circular. The green region indicates the shape of the plateau on top of the pillar. The last part is the whole EHT.

The movement of the EHTs is biologically defined. In Figure 2 an example of a trajectory of an EHT is shown. A certain base tension will push the pillars closer to each other even when they do not contract. Upon the base tension, the EHTs will contract and push the pillars to each other. The movement of the EHT is smooth, so no sudden changes should appear in the corresponding trajectory. In the videos, the EHT is stimulated with a 2Hz electronic pulse. This pulse will cause the EHTs to contract at 2Hz. Some of the EHTs will have a so-called ”double bump” in their trajectory. Although this bump is not biologically expected, it is present. This bump is an extra Contraction and appears when the tissue is not fully relaxed. In Figure 2 a ”double bump” trajectory is shown.

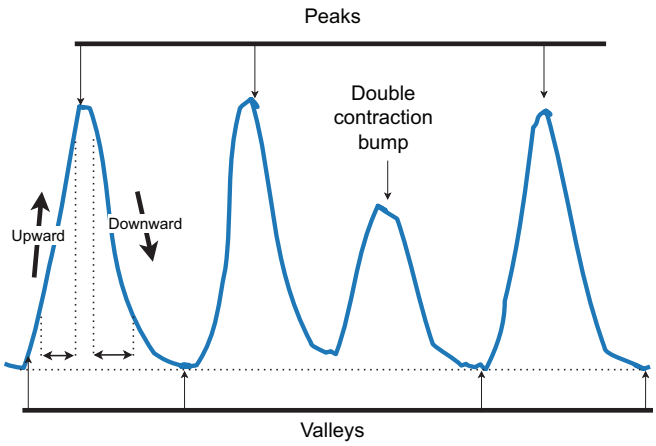

Fig. 2. An example of an EHTs displacement trajectory with so-called ”double bump” where the tissue contracts while not fully relaxed. Arrows mark the double bump.

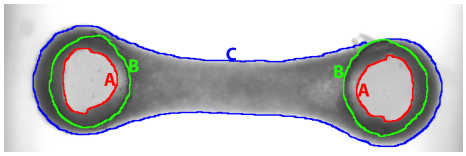

Fig. 3. EHT regions of interest. A. Inner area of the pillar. B. Area of the plate. C. EHT surface area.

IV. Methods

In this section, methods for tracking and segmentation of the EHT videos are presented. A overview of the total process is shown in Figure 4.

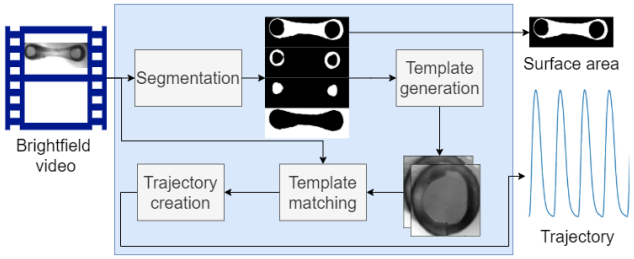

Fig. 4. Overview of the total EHT analysis algorithm.

A. Tracking

Finding the EHTs displacement consists of tracking the precise location of the pillars in each video frame.

1) Template matching: Template matching is a method for finding a (similar) region in a source image. Template matching compares a template image T to a region in a source image I by pixel intensities with normalized correlation coefficient(NCC) [33]. The corresponding formula's for NCC matching are found in Equation 1, Equation 2 and Equation 3. NCC puts out accumulator matrix R. Each value in the accumulator matrix represents a match score of the location

in the source image where one is an exact match, and zero is no match at all. We find the location of the best match by determining the highest score.

$$R(x,y) = \frac{\sum_{x',y'}(T'(x',y') * I'(x+x',y+y'))}{\sqrt{\sum_{x',y'} T'(x',y')^2 * \sum_{x',y'} I'(x+x',y+y')^2}} \quad (1)$$

$$T'(x',y') = I(x',y') - \frac{\sum_{x'',y''}(T(x'',y''))}{(w-h)} \quad (2)$$

$$I'(x+x',y+y') = I(x+x',y+y') - \frac{\sum_{x'',y''}(I(x'',y''))}{(w-h)} \quad (3)$$

In the videos of the EHTs different shapes of pillars, shapes of tissues, and lighting are present. To perform accurate template matching, a template is segmented for each individual video. This template is segmented on the first frame of the video. In subsection IV-B the proposed automatic pillar segmentation process is described. In Figure 5 an example of a tissue and its corresponding segmented template for the right pillar (Figure 7) is shown. The initial location or the location in the frame before is known. Also, from biology, it is known that the tissue moves within bounds. Therefore a region of interest around the known location is selected to compute the template matching algorithm. A significant reduction of computational costs is achieved by computing template matching only on the selected region. In Figure 6 the selected region of interest of Figure 5 is shown. The corresponding accumulator matrix is shown in Figure 8. Here the brighter the pixel is, the better the match. The pixel distance between the points is calculated to create a trajectory from the found points with Equation 4. Then the distance in pixels is converted to millimeters, and the distance between the pillars is subtracted with Equation 5. In Figure 9 a trajectory of an EHT gathered by template matching is shown.

$$T_{pixel} = \sqrt{(x_{left} - x_{right})^2 + (y_{left} - y_{right})^2} \quad (4)$$

$$T = T_{pixel} * \alpha - 3.2mm \quad (5)$$

Where α is the conversion factor from pixels to millimeters

2) Sub-pixel precision: Since the movement of the pillars is far below pixel level, subpixel precision is needed to gather usable results for the analysis of EHT videos. This subsection presents a new template/pattern matching approach with sub-pixel precision for tracking the movement of the pillar. The output of the NCC will give an accumulator matrix as is shown in Figure 8. In Figure 10 we zoom to the region of the maximum intensity. In this case, there is an exact match of the template. In Figure 10 the maximum score is in the center pixel. The pixel has a high score compared to its neighboring pixels. This is expected since shifting the template farther away from the correct location will lower the match score, and the middle pixel is an exact match. In Figure 11 the match lays between pixels, therefore, we see neighboring pixels that have a match score close to the maximum score. In Figure 11 we can

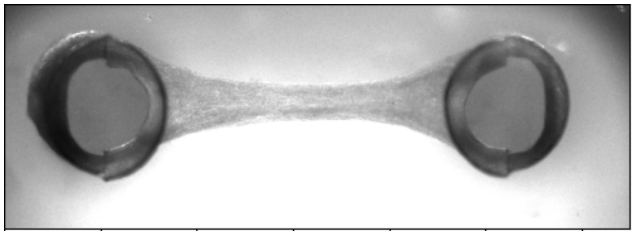

Fig. 5. brightfield of an EHT that is used as the input image for template matching.

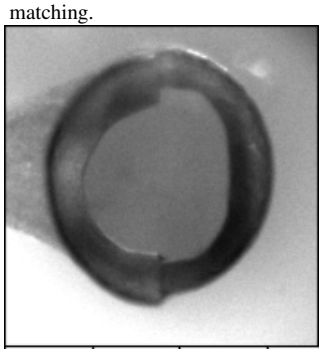

Fig. 6. Selected region of interest for template matching on a pillar of the EHT.

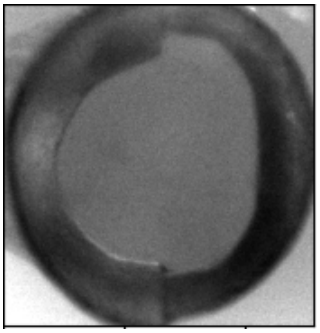

Fig. 7. Selected template for template matching a pillar of the EHT.

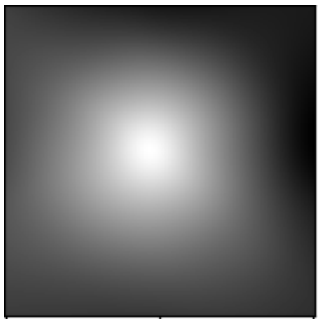

Fig. 8. Accumulator matrix of the template matching of the EHT

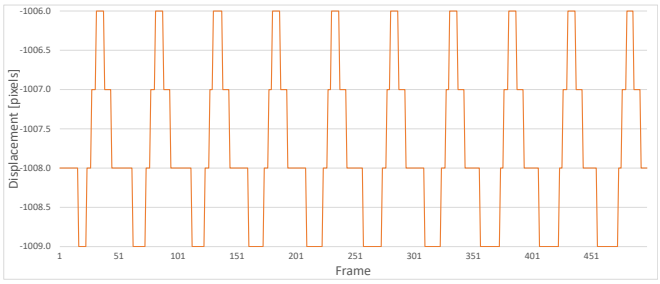

Fig. 9. Trajectory of an EHT tracked with template matching

see that the left neighbor pixel from the maximum pixel has a matching score close to the maximum. These values indicate that the exact match lies between the left neighbor pixel and the maximum score.

To find a match with subpixel precision, we use interpolation. Interpolation is a method that will estimate a function

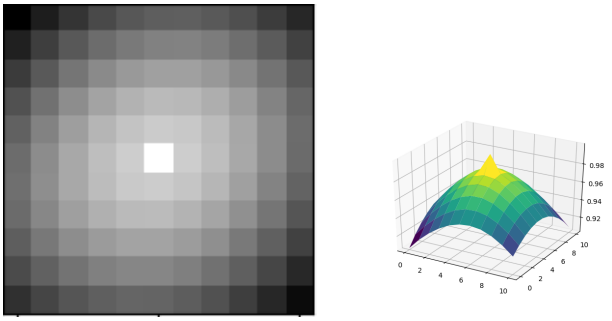

Fig. 10. Peak region result image of the template matching method exactly on pixel matched and its corresponding 3d projection

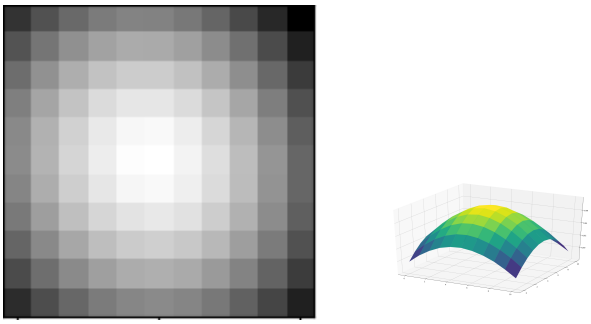

Fig. 11. Peak region result image of the template matching method exactly on pixel matched and its corresponding 3d projection

based on measured points. In Figure 11 it is observed that a linear function could not form the curves since the surface is expected to be smooth. A polynomial function can estimate the values between pixels and can find higher match scores between pixels. Bicubic spline is a common technique for obtaining smoothness in two-dimensional interpolation [34]. The bicubic spline can convert the peak region into a polynomial function. Bicubic spline outputs a function that gives a match score for a (decimal) location in the peak region and beyond. The local maximum of the function needs to be found to find the highest match score. Finding the local maximum is done by the Nelder–Mead algorithm. The Nelder–Mead algorithm is one of the best-known algorithms for multidimensional unconstrained optimization without derivatives [35]. The Nelder–Mead algorithm runs a converging process that computes test points to find the direction for the upcoming test points. It stops the converging process if the deviation between the test points is lower than a selected threshold. The location of the maximum match score at pixel level is already known and can therefore be a good starting point to find the local maximum of the function. The initial point is within a half-pixel distance of the maximum match score, and therefore, the computation of the Nelder–Mead algorithm is limited together with the threshold. The threshold for stopping the converging processes is 0.0001. No significant change to the outcome is present at this value since the level of noise in ETH videos is influencing the template matching algorithm significantly more. In Figure 12 the found location of Figure 11 shown. In

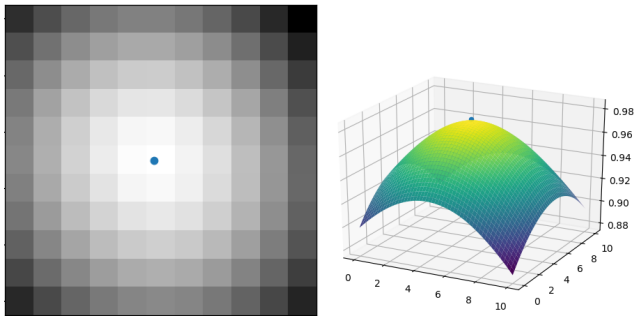

Fig. 12. Peak region result image by the template matching algorithm with annotated subpixel precision match and its corresponding interpolated 3d projection.

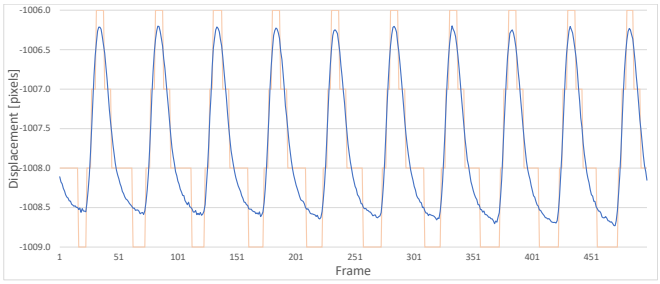

Fig. 13. Trajectory of an EHT tracked with template matching with subpixel precision.

Figure 12 the 3d surface shows interpolated version of the 3d surface in Figure 11. Now a whole trajectory can be made with the subpixel precision points found by the introduced method. An example of a trajectory can be found in Figure 13.

B. Segmentation by Deep learning

Segmentation methods based on deep learning approaches deploy neural networks to classify each pixel of an image belonging to one of the interest classes. The approaches do this by identifying different regions belonging to different (parts of) objects of interest. State-of-the-art approaches use convolutional networks that are trained using labeled examples. The label examples consist of layers that contain masks of the different regions. In DL, the most common approach is to apply a convolutional neural network (CNN) to analyze visual imagery [36].

1) Dataset: A dataset was created to provide a CNN with ground truth data of the EHT brightfield's. A set of 81 videos of about 500 brightfield frames per video are provided. Since manual segmenting is time-consuming, only four frames per video are segmented. For ten videos, seven frames per video are segmented. The dataset has 290 brightfield images with their corresponding labels. The labels consist of RGB images where white is the background, red is the inner pillar, green is the outer pillar, and blue is the tissue itself. An example of a label is shown in Figure 14. In the supplied dataset, there are two frame dimensions, 1414x532 pixels and 1476x472 pixels. Before we feed the data to the CNN, the data is resized to the

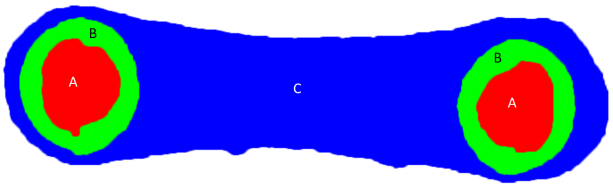

Fig. 14. Manual segmented label of the EHT dataset. A. Inner area of the pillar. B. Area of the plate. C. EHT surface area.

	U-net	Deeplabv3 with a ResNet-101 backbone
Loss function	Mean square error (MSE)	Mean square error (MSE)
Optimizer	Adaptive momentum (adam)	Adaptive momentum (adam)
Learn rate	0.05	0.01
Epochs	200	200
Batch size	1	3
Input layers	1	3
Output layers	4	4

TABLE I. Training settings for U-net and Deeplabv3 with a ResNet-101 backbone model on the EHT dataset

same dimensions. First, height and width are scaled equally to maintain the aspect ratio. Then padding is used to fill up space to the correct dimensions. Training a CNN is computationally costly. The dimension is resized to half of the maximum of the two input dimensions to decrease the computational cost. The output of the resize is an image with the dimension of 738x266 pixels. When training, the dataset is split into two parts. One for training, the other for testing. The test set should contain frames that the model has never trained on before. This, so actual performance on new videos can be measured. Since we use multiple frames per video, we must construct a test set of frames from videos not in the train set. Now the dataset is ready to use for training a model.

2) Training: A CNN contains convolutional layers that consist of a set of learnable filters (or kernels). These layers consist of artificial neurons. A CNN is set up to segment the regions of the EHTs. The input of the CNN is a brightfield frame of a EHT video. When fully working, the CNN will annotate the found shapes in mask matrixes. We initialize all parameters of the CNN network to start training it to give the correct outputs. Then the CNN is fed with frames of the EHT videos where the correct segmented output is known (labeled). We compare the output of the CNN with the known label by a cost function. The outcome of the comparison adjusts the CNN parameters. When the CNN learns enough samples, the CNN can segment on its own. We can test the precision of the CNN with a so-called test set. The test set contains labeled frames that are new to the CNN. We can give a score for precision to the CNN by comparing the outputs of the CNN with the labels. From the literature, we found two main CNN approaches for the segmentation of the EHT video frames, U-net [27] and Deeplabv3 with a ResNet-101 backbone [28]. Both approaches perform well in segmentation tasks where the U-net approach is developed for biomedical image segmentation.

3) U-net: For the training of the U-net model, we use the settings in Table I. The input layer of the model is adjusted from a three-layer RGB input to a one-layer grayscale input since the input is in grayscale.

4) Deeplabv3 with a ResNet-101 backbone: For the training of the Deeplab with a ResNet-101 backbone, we use the settings in Table I. The input layer of the model is in three-layer RGB format. Since the model uses the three layers further up in the model, the input layer cannot be changed directly. Therefore the grayscale input is converted to an RGB image before it is fed to the model.

5) Calculation the surface area: The output of the two proposed deep learning models is four mask images containing the three different regions of interest. The mask has the pixel value one if the pixel belongs to the class and zero if not. To calculate the surface area of the EHT, we calculate the surface of one pixel and multiply it by the sum of all activated pixels in the mask. In Equation 6 the formula for calculating the surface area is shown.

$$surface_{mm^2} = \alpha^2 * \sum I_{surfacemask} \quad (6)$$

Where α is the conversion factor from pixels to millimeters

C. Segmentation as input for template matching

Our template matching method needs templates of the two pillars to create a trajectory of the movement of the EHT. To supply the templates, manual or automatic segmentation can be used. This study proposes two deep learning models for the segmentation in chapter IV-B. To see if we can use the results of the deep learning models as an input to the template matching method, we make a comprehensive comparison between manual segmentation, segmentation by U-net, and segmentation by Deeplabv3 with a ResNet-101 backbone.

To create a template from a segmented mask, we can either pick the pillar's inner or outer part. We select a square around each pillar from the minimum and maximum bounds of its shape to create a template. The output of the deep learning models is expected to contain minor errors. These errors are unwanted activations that could lead to extra blobs in the segmentation masks. Because the blobs are expected to be smaller than the shape of the pillars, selecting the two pillars is done by picking the two larges shapes in the segmentation mask to generate the templates.

V. Verification

To the best of our knowledge, there are no ground truth data sets available for the segmentation and tracking of the EHT videos. Thus, a framework for verification is needed to evaluate the performance of the proposed approaches. This section proposes an experimental framework for verifying the quality of the segmentation and tracking results.

A. Tracking

For tracking the EHTs, subpixel precision is needed. For a human, it is impossible to create such precision with manual tracking. A human would produce too much error in the results, and therefore we can not use manual tracking to verify the tracking results. Furthermore, for a human, it would be time-consuming to place a marker on each video frame.

1) Simulation/Quantitative evaluation: First, a video of an EHT with marked black dots was used for a quantitative evaluation. The idea was to track the black dots to create the ground truth and then artificial remove the black dots from the video so the software could track it. We found that the black dots were not clean enough to track the pillars correctly in this experiment. Therefore a simulation video was used to create a ground truth.

To create a simulation video, we start with a background image. A video of an empty container of an EHT is filmed. This video is recorded on the same setup as the actual EHT videos and will give us the same noise level, illumination, and magnification. In figure Figure 15 a frame of the background video is shown. Now we can create a moving tissue on top of the background. An EHT is segmented from its video, cut in half, and placed on top of the background video. Since the pillars are moving inward, we need to move the two segmented parts to each other. The movement of the pillars is in subpixel precision. An affine transformation is used to translate the two parts to a sub-pixel precise location. By only moving the two parts inwards, we prevent a gap between the two parts showing the background. In Figure 16 a resulting frame of two segmented parts placed on top of the background is seen.

There is always noise in the actual EHT videos, and variabilities can be expected among videos. Therefore, random noise is added to the simulation video to test the robustness of the method. The background of the simulation video is recorded on the same setup and already contains noise. Therefore random noise is only added to the tissue. To test how well the method can resist noise, adjustments can be made to the amount of noise added. An example of a frame with 25% noise is shown in Figure 17. Different amplitudes, noise levels, and signals are generated to gather a wide variety of videos to get a good overview of the performance of the proposed method for tracking under different conditions. To measure the error from the ground truth, we calculate the mean absolute error (Equation 7) and mean square error (Equation 8).

$$MAD = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) \tag{7}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \tag{8}$$

2) Survey/Qualitative evaluation: Scientists with a biological background working with EHTs know how a certain EHT should move and how the corresponding trajectory should look. We can gather a good view of how the algorithm performs by surveying the trajectories produced by the tracking algorithm with these scientists. The survey contains a score from one to five, where one indicates that the trajectory is totally different from the expected and five is precisely as expected. Furthermore, the attendees to the survey should leave a comment to give clear reasoning about why the score was given.

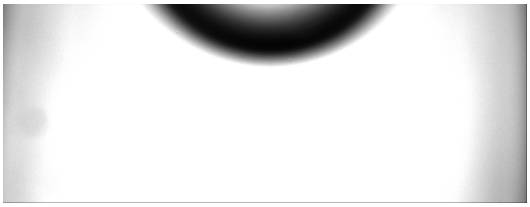

Fig. 15. Brightfield of an empty EHT container.

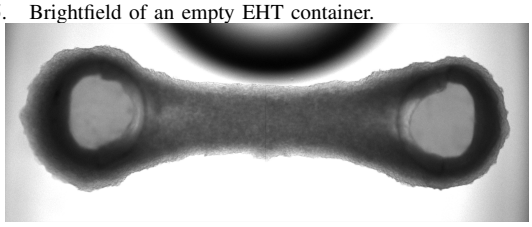

Fig. 16. Brightfield of an empty EHT container with two parts of an EHT placed on top.

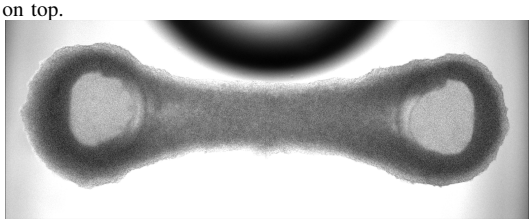

Fig. 17. Brightfield of an empty EHT container with two parts of an EHT placed on top and 25% noise added on the EHT.

3) Evaluation of EHT contractile properties: The tracking algorithm's output is a trajectory of the difference in distance over time of the pillars centers as a result of contraction of the tissue. An example of a trajectory is given in Figure 3. Biologically, The contractile properties of the EHTs are expected to behave according to:

1) The speed of contraction needs to be higher than the speed of relaxation.
2) There are no bumps when contracting.
3) During contraction or relaxation, the trajectory should never invert the direction.
4) There are no sudden spikes or block-shaped movement.
5) The contraction of the EHT should be according to the frequency of stimulation.

These behaviors can be tested on the trajectories produced by the template matching with sub-pixel precision method to verify the correctness of the tracking algorithm.

To test the behaviors, we first gather all the valleys and peaks. Then behavior 1 is tested by comparing the rise and fall speed. We measure rise speed by measuring the time it takes from the valley to the peak. For the fall speed we use the time from peak to valley. The last 10% at the top and bottom of the y axis of the contraction curve has a pretty different derivative.

We should exclude these parts in determining the rise and fall speed.

When an EHT is contracting, there should be no bumps while contracting. If there are no directional changes in the path from valley to peak, the trajectory complies to behavior 2. Behavior 2 is tested by checking each sample to be higher than its predecessor. For this test, the 10% top and bottom of the y axis are still left out for reliability since a small error in the trajectory could change its direction.

In behavior 3 it is stated that the direction cannot invert during contraction or relaxation. For the contraction, the check used in behavior 2 should be enough. As discussed in section I, when the EHT relaxes, a second contraction bump might be present. Therefore behavior 3 could only be tested partly. It would be good to measure the width and height of the contraction bumps since they are of interest to the scientists working with the EHTs. Measuring the width of the contraction bumps is done by measuring the time from the first direction change when falling, to the point that the same height is reached falling again from the contraction bump. The height of the contraction bump is measured by subtracting the peak of the contraction bump to its starting point.

An ideal trajectory does not contain any spikes or block-shaped movement. Since the camera's resolution is limited and noise is present, we expect that the results will not be perfect. Although we expect some noise and block-shaped movement, we can still test 4 by the derivative of the trajectory. Peaks and block-shaped movements have a high derivative. By placing a threshold on the derivative, we can pick out high peaks or blocky movements that significantly impact the trajectory.

The scientists stimulate the EHTs in the videos with a frequency of 2Hz. In behavior 5 it is stated that the EHTs will contract following this frequency. We check this behavior by determining the time between the peaks. When the frequency is significantly different from the stimulated frequency, the trajectory does not comply with behavior 5.

4) Comparison to state-of-the-art: To the best of our knowledge, the state-of-the-art for tracking ETHs is MUCLEMOTION [5]. We use the results of this algorithm for a comparison with the new proposed tracking algorithm. Furthermore, we use the simulation videos created in subsubsection V-A1 to compare both methods to the ground truth.

B. Segmentation

1) Manual segmentation: Although manual segmentation could contain small errors, it can still recognize outliers in the segmenting process. Intersection over Union (IoU) is used to compare the manual segmented results to the proposed method.

$$IoU = \frac{AreaofOverlap}{AreaofUnion} \tag{9}$$

VI. Results

In this section, results are presented of the proposed tracking and segmenting algorithms and their verification methods.

A. Template/pattern matching with Sub-pixel precision

1) Simulation/Quantitative evaluation: In Figure 18 and Figure 19 the results of the proposed tracking method to the simulation videos with variation levels of noise are shown. Here, we can observe that the error grows when noise becomes higher. The algorithm is stable until about 25% noise is added to the tissue in the simulation video. Above the 25%, the error is too high to gather stable results. With low noise added, the algorithm performs well under 0.1 pixel precision.

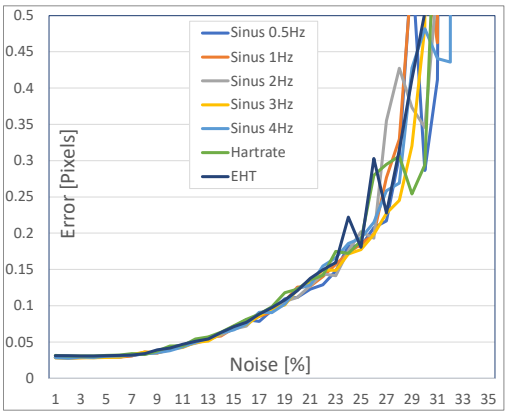

Fig. 18. Mean absolute error results of the simulated videos of different signals tracked by the proposed template/pattern matching with sub-pixel precision method compared to the ground truth.

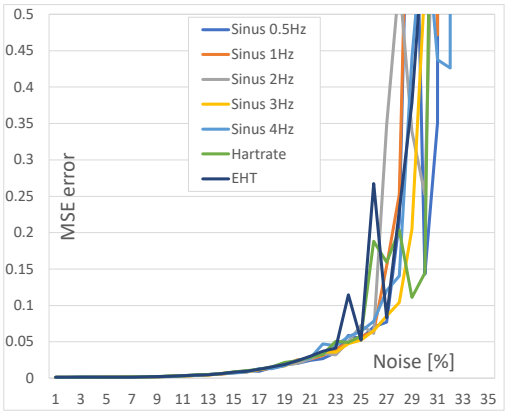

Fig. 19. Mean square error results of the simulated videos of different signals tracked by the proposed template/pattern matching with sub-pixel precision method compared to the ground truth.

2) Survey/Qualitative evaluation: A specialist who works with EHT videos has taken a look at the results of a dataset of 81 EHT videos. In Figure 20 the results of the survey are shown.

3) Evaluation of EHT contractile properties: In Figure 21 it can be seen that most of the videos contract faster than they relax. There are three videos where the contraction is slower than the relaxation. In Figure 22 it can be seen that the frequency in most of the videos is very close to $2Hz$ with a maximum outlier of $2.25Hz$. In Figure 23 it can be seen that

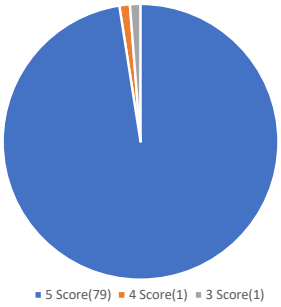

Fig. 20. Scores of the tracked trajectories scored by a specialist from one to five, where one is totally different from the expected and five is precisely as expected.

most of the videos do not have rise direction change errors. In Figure 24 it can be seen that when the displacement is low, the amount of rise direction change errors becomes higher and when the displacement is higher than about $0.05mm$ no errors are present. In Figure 25 it can be seen that most videos have a maximum displacement speed between $0 - 150mm/s$ and that there are four outliers. In Figure 26 we can see a clear correlation between the maximum displacement and the maximum speed of the trajectory. In Figure 27 it can be seen that most videos have a displacement between $0 - 10mm$ and that there are four outliers. In Figure 28 we can see that the four outliers with a higher displacement also have a higher surface area.

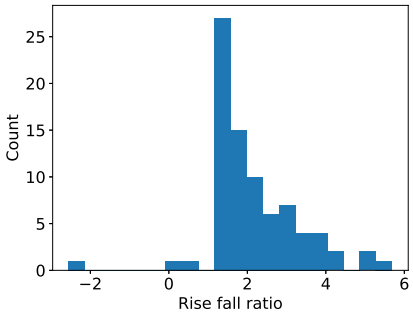

Fig. 21. Histogram of the fall versus rise speed ratio of the 81 EHT videos.

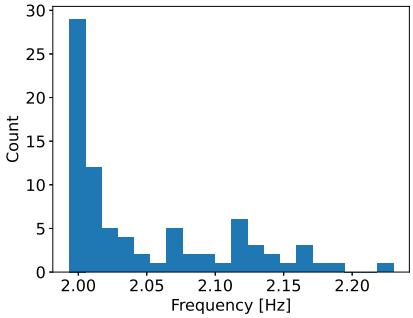

Fig. 22. Histogram of the frequency of the 81 EHT videos.

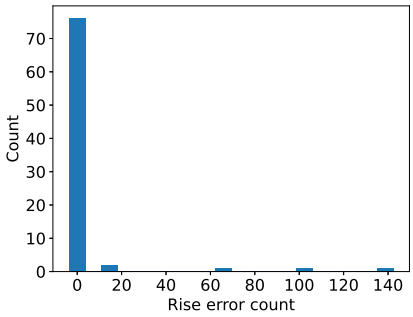

Fig. 23. Histogram of the rise direction error counts in the trajectory of the 81 EHT videos.

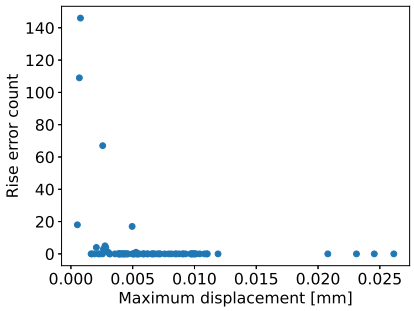

Fig. 24. Maximum displacement versus rise direction error counts of the trajectory of the 81 EHT videos.

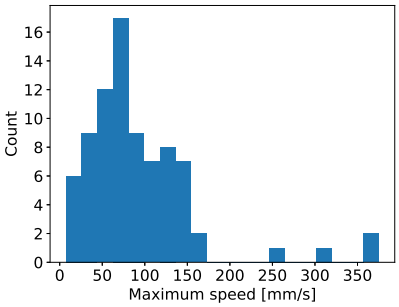

Fig. 25. Histogram of the maximum displacement speed of the trajectory of the 81 EHT videos.

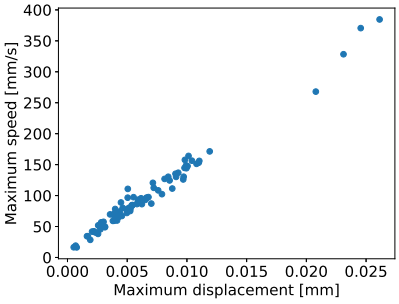

Fig. 26. Maximum displacement versus the maximum displacement speed of the trajectory of the 81 EHT videos.

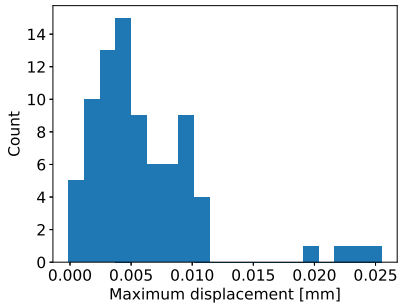

Fig. 27. Histogram of the maximum displacement of the trajectory of the 81 EHT videos.

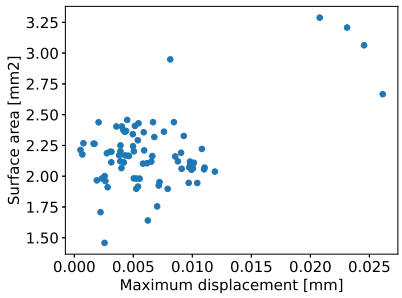

Fig. 28. Maximum displacement versus the surface area of the 81 EHT videos.

4) Comparison to state-of-the-art: In Figure 29 it can be seen that both algorithms have found the trajectory. MUSCLEMOTION starts from a high point, and then it lowers more than 20 pixels. In Figure 30 MUSCLEMOTION detects the same trajectory, but in the middle of the y-axis, it suddenly changes direction. In Figure 31 MUSCLEMOTION is compared with template matching with sub-pixel precision and with the ground truth. Both algorithms are close to the ground truth. In Figure 32 we see that although a noisy video template matching with sub-pixel precision still finds the ground truth trajectory. In appendix A more comparisons are shown.

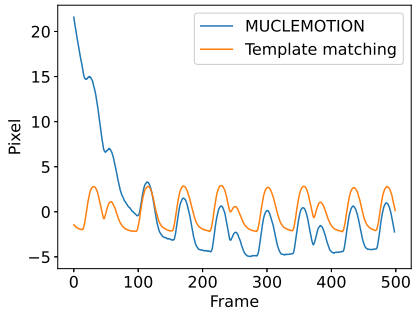

Fig. 29. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

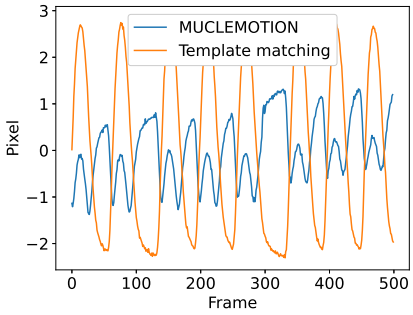

Fig. 30. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

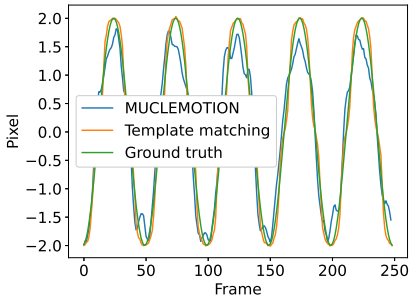

Fig. 31. Comparison of a trajectory from a simulation video of a 2Hz sine with a 2-pixel amplitude and 10% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

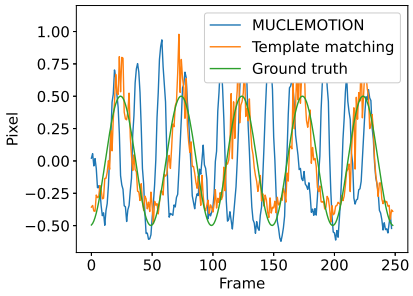

Fig. 32. Comparison of a trajectory from a simulation video of a 2Hz sine with a 0.5-pixel amplitude and 20% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

B. Segmentation by deep learning

The result of a Deeplabv3 with a ResNet-101 backbone and a U-net prediction is shown in Figure 33. In Table II the results of the IoU of the test set are shown.

Model	Outer	Tissue	Outer Pillar	Inner Pillar
deeplabv3 resnet101[28]	96.8%	96.0%	90.1%	91.2%
Unet[27]	97.3%	97.1%	91.3%	92.5%

TABLE II. Intersection over Union (IoU) results of the Deeplabv3 with a ResNet-101 backbone and U-net prediction of the background, inner pillar, outer pillar and tissue classes

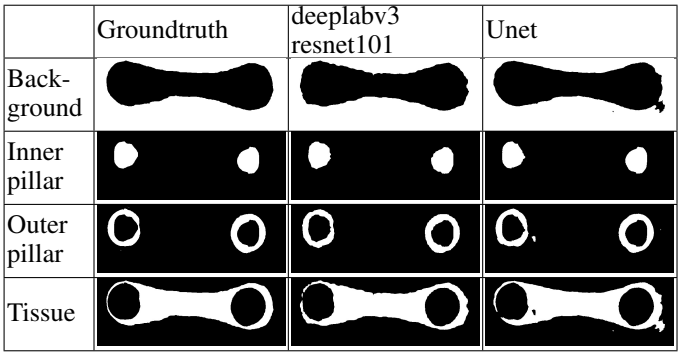

Fig. 33. Segmentation ground truth, Deeplabv3 with a ResNet-101 backbone prediction and Unet prediction of the background, inner pillar, outer pillar and tissue classes

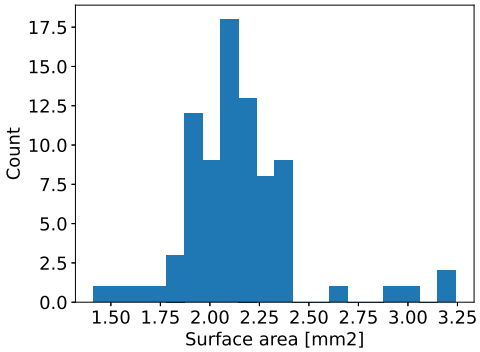

Fig. 34. Histogram of the surface areas of 81 EHT videos.

1) Calculation the surface area: The calculation of the surface area relies on the performance of the segmentation. Therefore, the results of the segmentation in paragraph VI-B can be directly related to the performance of the surface area calculation. In Figure 34 a histogram of the surface areas of the base set of 81 videos is shown. We can see a clear peak around $2mm^2$.

C. Segmentation as input for template matching

1) Comprehensive comparison of templates: In Figure 35 and Figure 36 the results of the comprehensive comparison of different templates are shown. We can see that the various trajectories have different base distances since they are not on the same height. Although the different base distances, the trajectories look very similar. Overall, the outer pillar templates have a smoother trajectory than the inner pillar template trajectories. In some cases, the base distance of two or more methods is the same as shown in Figure 36, but this varies between videos.

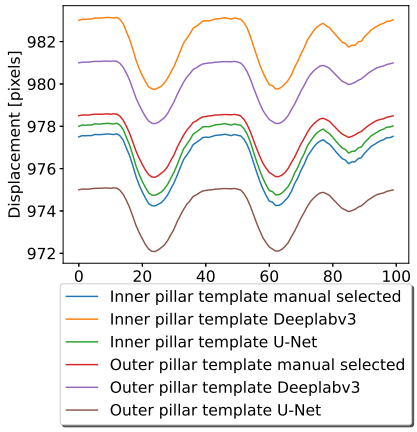

Fig. 35. Comparison between templates. Inner and Outer part of a pillar as template manual selected or by segmentation by U-net and Deeplabv3

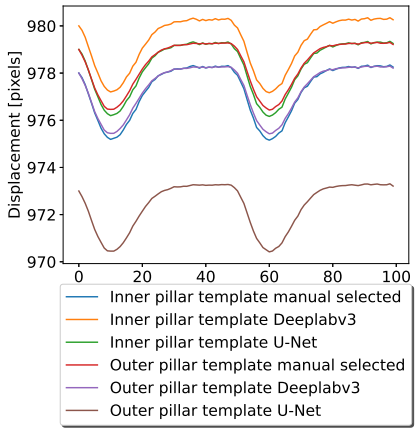

Fig. 36. Comparison between templates. Inner and Outer part of a pillar as template manual selected or by segmentation by U-net and Deeplabv3

D. Computational speed

The computation speed of the algorithms is measured and differ by the computers performance. In Table III the computation speed results of the proposed methods are shown on a high end desktop and a laptop.

	laptop		desktop	
	CPU	GPU	CPU	GPU
	i7 6700HQ	GTX960m	I7 10700	RTX3090
Video loading	4.7s	x	1.5s	x
U-Net model load	0.21s	0.33s	0.13s	0.15s
U-Net segmentation	2.15s	2.13s	1.0s	1.02s
Deeplabv3 model load	2.31s	5.34s	1.19s	2.19s
Deeplabv3 segmentation	4.91s	5.17s	2.43s	2.47s
Inner pillar template matching	14.1s	x	6.2s	x
Outer pillar template matching	16.8s	x	7.2s	x

TABLE III. speed results of loading, segmentation and template matching

VII. Discussion

A. Template/pattern matching with Sub-pixel precision

1) Simulation/Quantitative evaluation: A great addition to the method would be to measure the level of noise present in a video before the analysis. This measurement can then estimate the precision of the output trajectory with the found results of the simulation.

2) Survey/Qualitative evaluation: The survey shows that in 98% the tracking algorithm produces a precisely as expected trajectory. One video had a score of 3 because one of the pillar centers was moving up and down. This movement was due to the printing on the screen. The trajectory was correct. One video had a score of 4 because the point was not correctly in the center of the pillar during the whole video. This misalignment had to do with a not correctly manual segmented template. When we leaf the two videos out, we see that the algorithm is 100% successful in producing a precisely as expected trajectory. This result shows that the algorithm is stable and of great use for the scientist working with the EHTs.

3) Evaluation of EHT contractile properties: Most EHT videos comply too the property evaluation of the rise speed that should always be higher than the fall speed. In the video where the ratio of contraction speed versus relaxation speed was negative, the trajectory was noisy, and the displacement was very low, causing the detection algorithm to select the wrong peaks resulting in a negative result. In the two videos where the contraction speed versus relaxation speed ratio was below one, a double contraction bump by the tissue was detected as another contraction. This bump causes the calculation of the fall speed to be higher since it only takes the top part of relaxation into account, resulting in a ratio below one. The verification algorithm suffers when the trajectory is noisy or contains small double bumps that are not detected. Although the verification algorithm suffers under these circumstances, the tracking algorithm is still correct after manual checking. Some trajectories like Figure 37 have rise direction errors. Here the direction changes while rising. In Figure 26 we can see that this happens especially when the pixel displacement is low. This is as expected. From the simulation in paragraph VI-A3, we can conclude that there is an absolute error in template matching with sub-pixel precision. When the pixel displacement becomes lower, the relative error rises, making it more likely that a rising error happens. A low pass filter on the trajectories could solve this partly. When tissue has no displacement, it means that physiologically the tissue is dead. A trajectory of a tissue that has low displacement is of less relevance to the scientists since it is almost dead. The rise errors in the videos are caused by the relative error in the tracking algorithm but only appear in less relevant EHT videos. The relative error problem also counts for the behavior that there are no sudden spikes or block shape movement. In trajectories with a very low pixel displacement, more errors will appear. In parts of trajectories with less movement, for example, the lower parts of the contraction, more errors will appear.

Since electronic pulses stimulate the frequency of contraction, the measured frequency of contraction in the EHT videos should be the same. We see that most videos are very close to the stimulated $2Hz$. Some videos have a slightly higher frequency. In the trajectories of these videos, we see that double contraction bumps are present, causing the contraction to be irregular. In Figure 37 a trajectory distorted by double contraction bumps is seen. Here the maximum frequency is $2.63Hz$ and the minimum $1.59Hz$ the average is $2.12Hz$. It is seen that every time a double contraction bump has appeared, the following two contractions are very close to each other, causing a higher frequency. The double contraction bump makes the minimum and maximum frequency diverge more. When we look at the average frequency, most videos still comply with a small margin to the stimulated $2Hz$.

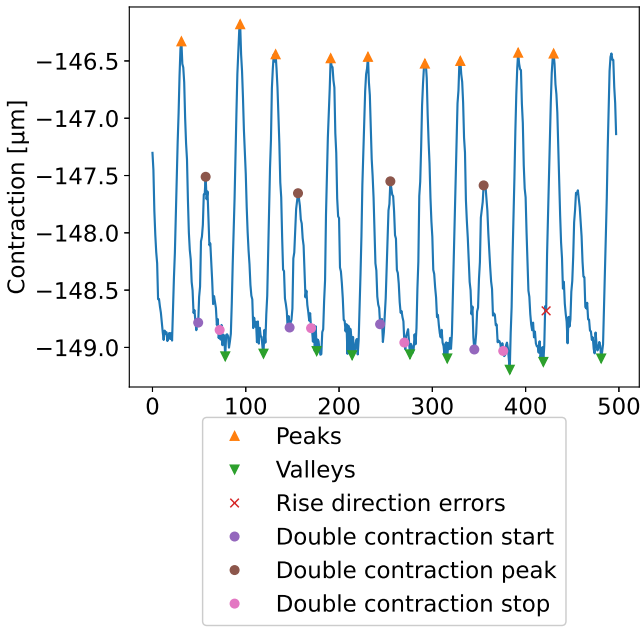

Fig. 37. Trajectory of a EHT with double bumps and low displacement.

4) Comparison to state-of-the-art: In comparing MUSCLEMOTION and the proposed method, we compared actual EHT videos and simulation videos. The simulation looks artificial, and in the middle, where the two parts of the tissue overlap, a stripe appears. Because the template matching with sub-pixel precision method uses a region of interest, the stripe part will not be considered. If it appeared on a region of interest, it still would deliver a low match since it is not expected that a template also contains a stripe. The MUSCLEMOTION method uses subtraction of a reference frame to the current frame to measure. MUSCLEMOTION could measure the differences at the stripe and therefore we should take caution when taking a conclusion from the simulation results.

In Figure 29 we see that MUSCLEMOTION finds the same curves, but the base tension changes from the start. This tension change is not expected since the tissue should return to its base tension when relaxing. In Figure 30 we see that MUSCLEMOTION changes direction halfway contraction and relaxation. From the biological defines, we know that this

should not happen. In Figure 31 and Figure 32 a comparison is made with simulation. As discussed, this comparison can suffer from the artificial effects of the stripe. In Figure 31 we see that both methods are close to the ground truth where MUSCLEMOTION has more distortion in the trajectory. In Figure 32 we see that the distortion is high. The 25% of noise and the relatively low displacement cause this distortion. MUSCLEMOTION could not find any trajectory that looks like the ground truth, but template matching with sub-pixel precision does. Summarizing the results of the comparison, we see that template matching with sub-pixel precision outperforms MUSCLEMOTION.

B. Segmentation by Deep learning

From the results, we found out that the U-net implementation scores the best in the IoU test. Although IoU scores higher, the visual results show that the U-net implementation shows more strange blobs in unexpected places (Figure 33). When selecting the pillar templates for the tracking algorithm, blobs may cause inaccuracies. It is better to have as few as possible blobs in the output. The Deeplabv3 implementation would be the best to pick for the segmenting task since the Deeplabv3 implementation scores very close to the U-net implementation and suffers less from the appearance of blobs.

1) Determining the surface area: Since the result of the segmentation directly represent the results of the surface area calculation. We can see that with a small error, the segmentation can determine the surface area of the EHT. In Figure 34 we see a nice spread of the surface areas between the data set of 81 videos. The four videos with a relatively high surface area also have a high displacement in their trajectory, explaining their outlying size.

C. Segmentation as input for template matching

1) Comprehensive comparison of templates: The resulting trajectories stay more or less the same between templates. The outer pillar produces the smoothest trajectories. This probably has to do with more variations and edges in the outer pillars template, resulting in more errors when a template is not in the correct location. The base tension differs between methods but also between methods and videos. The base tension that differs between methods has to do with the center of a template, which can differ a few pixels per method. Boxes are drawn around the found pillar shapes and determine the centers. The difference between videos can be caused by the shape of the silicon pillars that can differ by their production process, and the shape of the tissue can slightly alter the pillar shape. Although the initial position slightly changes from the manual selected template, both Deeplabv3 and U-Net succeed in selecting the correct template. As discussed in paragraph VII-B the Deeplabv3 implementation show fewer blobs in its results, making it better for displaying and creating templates.

D. Computational speed

The computational speed results are in the range of seconds resulting in high throughput. Scientists can now analyze whole batches of EHT videos automatically in a reasonable time. Moreover, since the algorithm is fully automatic, no manual work is needed.

E. Practical problem

The template matching with sub-pixel precision produces trajectories that are of great use for the scientists who work with EHTs. In developing this method, the scientists started using it immediately since it was of such gain to the methods they used before for gathering trajectories. Combined with the automatically picking of the template by segmentation, this method solved the analyzing problem of the scientists who work with EHTs. The extra features gathered by the biological checks are also of great relevance for scientists who work with ETHs. In Figure 26 we see a correlation of the movement speed in the trajectory versus the maximum displacement of the EHT. This correlation is expected since the displacement and the contraction speed are both related to the contraction force.

This tool will allow to check the the Frank-Starling law of the heart [37] which relates the contraction force to the volume of the tissue. To check the Frank-Starling law of the heart a different platform is needed.

VIII. Conclusion

The proposed method for tracking and segmentation is a huge step forward in the assessment of EHTs. Scientists who work with EHTs can now do an easy, quick, and fully automatic analysis with highly accurate results. The results given by the algorithm are proven to be robust and are tested by the verification framework. The additional results of the biological checks also prove great potential in the information that can be subtracted from the proposed methods. The method will help in the assessment of the EHTs for biological and drug discovery studies.

The deep learning-based segmentation can provide templates for the template matching with the sub-pixel precision method. We conclude that the Deeplabv3 with a ResNet-101 backbone implementation works the best and that the template of the pillar plate creates the smoothest trajectories. The high accuracy of the deep learning-based segmentation allows the use of the segmented tissue shape to calculate the surface area of the EHT.

The quantitative evaluation results by simulation, evaluate that template matching with sub-pixel precision is accurate when noise levels are within the limit. From the qualitative evaluation in the form of a survey under the scientists who work with EHTs, we can conclude that the resulting trajectories are as expected. Although some outliers, we can conclude that the results comply with the defined biological behaviors. In comparison to the state-of-the-art, we conclude that the results look similar but that the proposed method outperforms the state-of-the-art.

Platforms for EHTs that are not trained by the deep learning segmentation model can still be tracked by the proposed tracking method by manually selecting the pillars or anchors.

This makes the tracking method very flexible for experiments with different platforms.

Although not explicitly developed, we can use this method in many more fields. For example, the study of contraction of single cells where an edge could be picked as a template. In non-biological fields, for example, in a vibration or shock tests of electronic devices the proposed method can be used to study the precise movement of parts.

IX. Python Package

A python package is created for tracking with sub-pixel precision to provide scientists with an easy-to-use library. This library does not contain the segmentation part since the segmentation relies on a specifically trained model, which will not work on slightly different images. Instead, the scientists can provide their own templates.

The python package can be found on www.github.com/dkeekstra/sp_templatematching.

X. Acknowledgements

I want to thank Nicola Strisciuglio for helping and guiding me during the study. I want to thank José Rivera Arbelaez and Marcelo Catarino Ribeiro for their support on the biology-related aspects of the project and for providing data of the EHTs.

REFERENCES

[1] A. Timmis, N. Townsend, C. P. Gale, A. Torbica, M. Lettino, S. E. Petersen, E. A. Mossialos, A. P. Maggioni, D. Kazakiewicz, H. T. May, D. De Smedt, M. Flather, L. Zuhlke, J. F. Beltrame, R. Huculeci, L. Tavazzi, G. Hindricks, J. Bax, B. Casadei, S. Achenbach, L. Wright, P. Vardas, and E. S. of Cardiology, "European Society of Cardiology: Cardiovascular Disease Statistics 2019," *European Heart Journal*, vol. 41, no. 1, pp. 12–85, 12 2019. [Online]. Available: https://doi.org/10.1093/eurheartj/ehz859

[2] B. V. Reamy, P. M. Williams, and D. P. Kuckel, "Prevention of cardiovascular disease," *Primary Care: Clinics in Office Practice*, vol. 45, no. 1, pp. 25–44, 2018, cardiovascular Disease. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0095454317301380

[3] J. M. Stein, C. L. Mummery, and M. Bellin, "Engineered models of the human heart: directions and challenges," *Stem Cell Reports*, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213671120304616

[4] S. Schaaf, A. Shibamiya, M. Mewe, A. Eder, A. Stöhr, M. N. Hirt, T. Rau, W.-H. Zimmermann, L. Conradi, T. Eschenhagen, and A. Hansen, "Human engineered heart tissue as a versatile tool in basic research and preclinical toxicology," *PLOS ONE*, vol. 6, no. 10, pp. 1–11, 10 2011. [Online]. Available: https://doi.org/10.1371/journal.pone.0026397

[5] L. Sala, B. J. van Meer, L. G. Tertoolen, J. Bakkers, M. Bellin, R. P. Davis, C. Denning, M. A. Dieben, T. Eschenhagen, E. Giacomelli, C. Grandela, A. Hansen, E. R. Holman, M. R. Jongbloed, S. M. Kamel, C. D. Koopman, Q. Lachaud, I. Mannhardt, M. P. Mol, D. Mosqueira, V. V. Orlova, R. Passier, M. C. Ribeiro, U. Saleem, G. L. Smith, F. L. Burton, and C. L. Mummery, "Musclemotion," *Circulation Research*, vol. 122, no. 3, pp. e5–e16, 2018. [Online]. Available: https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.117.312067

[6] R. M. Jiang, D. Crookes, N. Luo, and M. W. Davidson, "Live-cell tracking using sift features in dic microscopic videos," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, pp. 2219–2228, 2010.

[7] T. Hayakawa, T. Kunihiro, S. Dowaki, H. Uno, E. Matsui, M. Uchida, S. Kobayashi, A. Yasuda, T. Shimizu, and T. Okano, "Noninvasive evaluation of contractile behavior of cardiomyocyte monolayers based on motion vector analysis," *Tissue Engineering Part C: Methods*, vol. 18, no. 1, pp. 21–32, 2012, pMID: 21851323. [Online]. Available: https://doi.org/10.1089/ten.tec.2011.0273

[8] T. Hayakawa, T. Kunihiro, T. Ando, S. Kobayashi, E. Matsui, H. Yada, Y. Kanda, J. Kurokawa, and T. Furukawa, "Image-based evaluation of contraction–relaxation kinetics of human-induced pluripotent stem cell-derived cardiomyocytes: Correlation and complementarity with extracellular electrophysiology," *Journal of Molecular and Cellular Cardiology*, vol. 77, pp. 178–191, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022282814002910

[9] M. Ghanbari, "The cross-search algorithm for motion estimation," *IEEE Transactions on Communications - TCOM*, vol. 38, 07 1990.

[10] E. Ibarra, R. Medina, V. Morocho, and P. Vanegas, "Optical flow as a tool for cardiac motion estimation," in *2015 Asia-Pacific Conference on Computer Aided System Engineering*, 2015, pp. 173–178.

[11] P. Torkashvand, H. Behnam, and Z. A. Sani, "Modified optical flow technique for cardiac motions analysis in echocardiography images," *J Med Signals Sens*, vol. 2, no. 3, pp. 121–127, Jul 2012.

[12] L. Thomas and J. Gehrig, "Multi-template matching: a versatile tool for object-localization in microscopy images," *BMC Bioinformatics*, vol. 21, 02 2020.

[13] P. R. Wankhede and K. B. Khanchandani, "Optic disc detection using template matching based on color plane histograms," in *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, 2016, pp. 278–281.

[14] M. Sun, J. Xiao, E. G. Lim, B. Zhang, and Y. Zhao, "Fast template matching and update for video object tracking and segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 791–10 799.

[15] V. R. Nafisi, M. H. Moradi, and M. H. Nasr-Esfahani, "A template matching algorithm for sperm tracking and classification," *Physiological measurement*, vol. 26, no. 5, p. 639, 2005.

[16] M. Rahimzadeh, A. Attar *et al.*, "Sperm detection and tracking in phase-contrast microscopy image sequences using deep learning and modified csr-dcf," *arXiv preprint arXiv:2002.04034*, 2020.

[17] A. Lukežič, T. Vojíř, L. Č. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter tracker with channel and spatial reliability," *International Journal of Computer Vision*, vol. 126, no. 7, pp. 671–688, 2018.

[18] A. Dutta, A. Mondal, N. Dey, S. Sen, L. Moraru, and A. E. Hassanien, "Vision tracking: A survey of the state-of-the-art," *SN Computer Science*, vol. 1, no. 1, pp. 1–19, 2020.

[19] M. V. Caya, D. Padilla, G. Ombay, and A. J. Hernandez, "Detection and counting of red blood cells in human urine using canny edge detection and circle hough transform algorithms," in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, 2019, pp. 1–5.

[20] S. M. Mazalan, N. H. Mahmood, and M. A. A. Razak, "Automated red blood cells counting in peripheral blood smear image using circular hough transform," in *2013 1st International Conference on Artificial Intelligence, Modelling and Simulation*, 2013, pp. 320–324.

[21] D. J. Kerbyson and T. J. Atherton, "Circle detection using hough transform filters," in *Fifth International Conference on Image Processing and its Applications, 1995.*, 1995, pp. 370–374.

[22] S. Sekhar, W. Al-Nuaimy, and A. K. Nandi, "Automated localisation of retinal optic disk using hough transform," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2008, pp. 1577–1580.

[23] T. Tran, O.-H. Kwon, K.-R. Kwon, S.-H. Lee, and K.-W. Kang, "Blood cell images segmentation using deep learning semantic segmentation," in *2018 IEEE International Conference on Electronics and Communication Engineering (ICECE)*, 2018, pp. 13–16.

[24] M. Imran Razzak and S. Naz, "Microscopic blood smear segmentation and classification using deep contour aware cnn and extreme machine learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[25] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1562–1573, 2018.

[26] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[28] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *CoRR*, vol. abs/1706.05587, 2017. [Online]. Available: http://arxiv.org/abs/1706.05587

[29] J. Li, H. Mu, and W. Xu, "A method of using digital image processing for edge detection of red blood cells," in *Sensors and Transducers Journal*, vol. 159, no. 11, 2011.

[30] S. Suwanmanee, S. Chatpun, and P. Cabrales, "Comparison of video image edge detection operators on red blood cells in microvasculature," in *The 6th 2013 Biomedical Engineering International Conference*, 2013, pp. 1–4.

[31] M. Mohamed and B. Far, "An enhanced threshold based technique for white blood cells nuclei automatic segmentation," in *2012 IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2012, pp. 202–207.

[32] A. Khashman and E. Al-Zgoul, "Image segmentation of blood cells in leukemia patients," in *Proceedings of the 4th WSEAS International Conference on Computer Engineering and Applications*, ser. CEA'10. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2010, p. 104–109.

[33] A. Kaehler and G. Bradski, *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*. O'Reilly Media, 2016. [Online]. Available: https://books.google.nl/books?id=SKy3DQAAQBAJ

[34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical recipes in c," pp. 123–128, 1988.

[35] S. Singer and J. Nelder, "Nelder-Mead algorithm," *Scholarpedia*, vol. 4, no. 7, p. 2928, 2009, revision #91557.

[36] M. Valueva, N. Nagornov, P. Lyakhov, G. Valuev, and N. Chervyakov, "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation," *Mathematics and Computers in Simulation*, vol. 177, pp. 232–243, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378475420301580

[37] J. P. Konhilas, T. C. Irving, and P. P. De Tombe, "Frank-starling law of the heart and the cellular mechanisms of length-dependent activation," *Pflugers Archiv*, vol. 445, no. 3, pp. 305–310, 2002.

APPENDIX

A. Results comparison MUSCLEMOTION

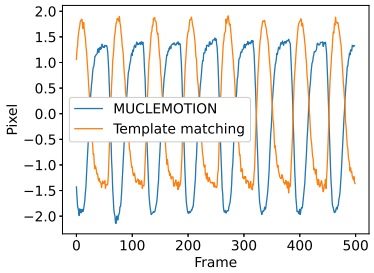

Fig. 38. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

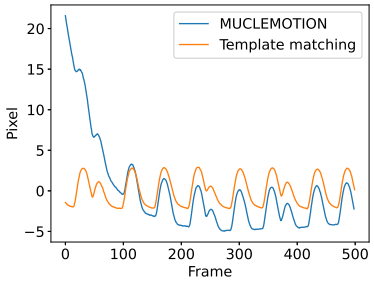

Fig. 39. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

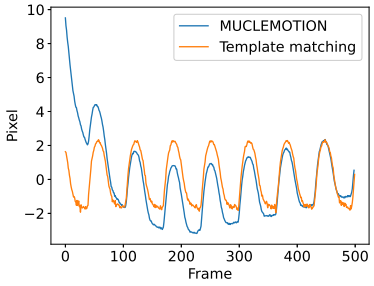

Fig. 40. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

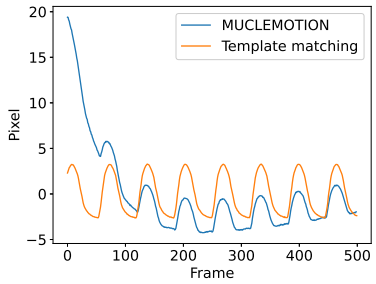

Fig. 41. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

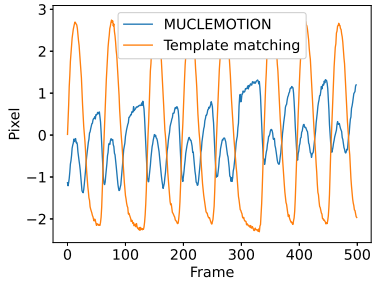

Fig. 42. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

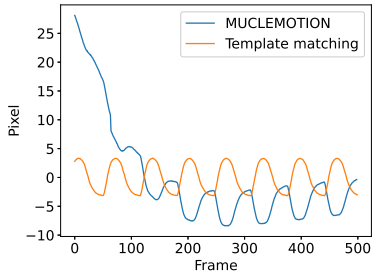

Fig. 43. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

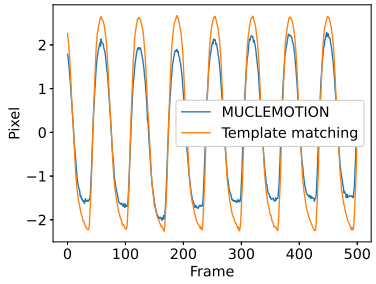

Fig. 44. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

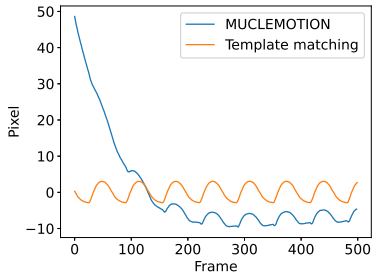

Fig. 45. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

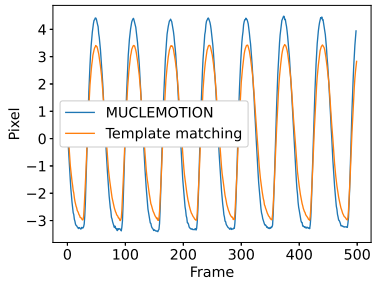

Fig. 46. Comparison of a trajectory from a EHT video, MUSCLEMOTION versus template matching with sub-pixel precision.

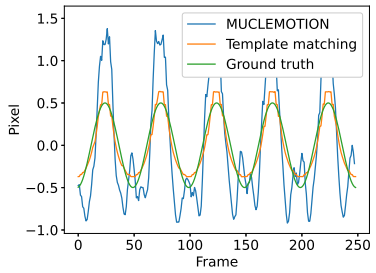

Fig. 47. Comparison of a trajectory from a simulation video of a 2Hz sine with a 0.5-pixel amplitude and 1% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

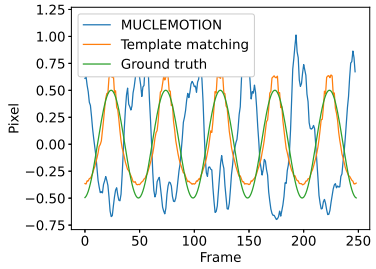

Fig. 48. Comparison of a trajectory from a simulation video of a 2Hz sine with a 0.5-pixel amplitude and 5% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

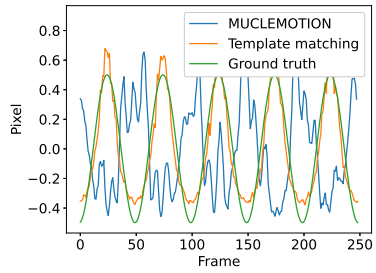

Fig. 49. Comparison of a trajectory from a simulation video of a 2Hz sine with a 0.5-pixel amplitude and 10% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

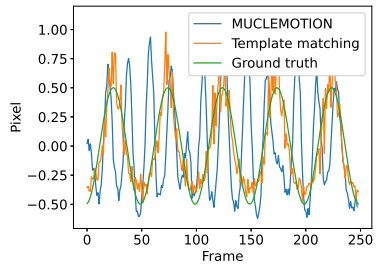

Fig. 50. Comparison of a trajectory from a simulation video of a 2Hz sine with a 0.5-pixel amplitude and 20% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

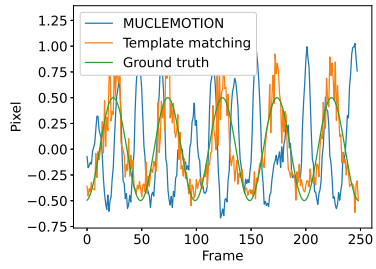

Fig. 51. Comparison of a trajectory from a simulation video of a 2Hz sine with a 0.5-pixel amplitude and 25% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

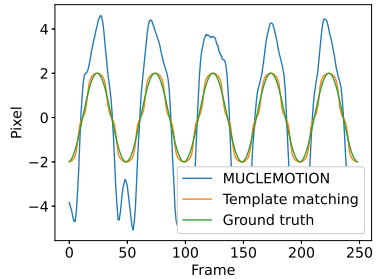

Fig. 52. Comparison of a trajectory from a simulation video of a 2Hz sine with a 2-pixel amplitude and 1% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

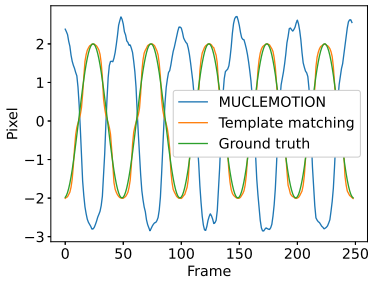

Fig. 53. Comparison of a trajectory from a simulation video of a 2Hz sine with a 2-pixel amplitude and 5% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

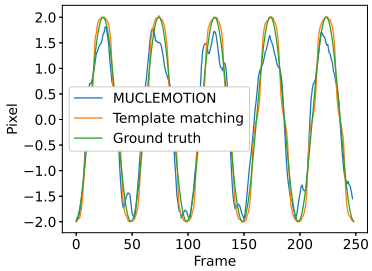

Fig. 54. Comparison of a trajectory from a simulation video of a 2Hz sine with a 2-pixel amplitude and 10% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

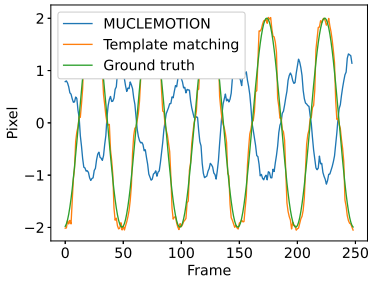

Fig. 55. Comparison of a trajectory from a simulation video of a 2Hz sine with a 2-pixel amplitude and 20% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

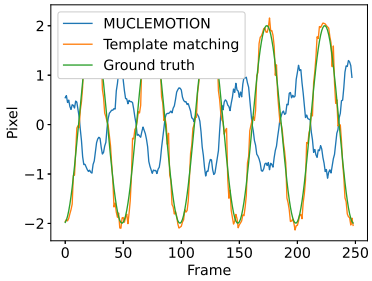

Fig. 56. Comparison of a trajectory from a simulation video of a 2Hz sine with a 2-pixel amplitude and 25% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

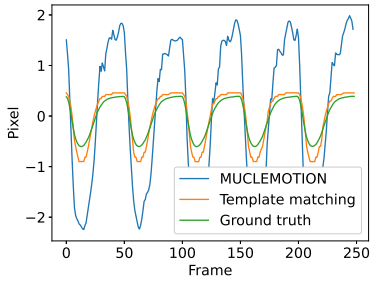

Fig. 57. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 0.5-pixel amplitude and 1% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

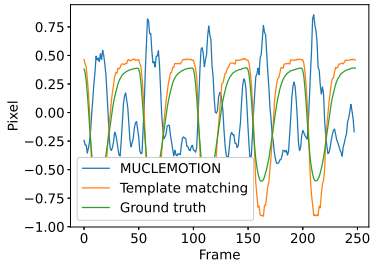

Fig. 58. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 0.5-pixel amplitude and 5% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

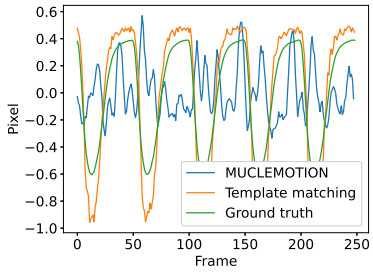

Fig. 59. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 0.5-pixel amplitude and 10% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

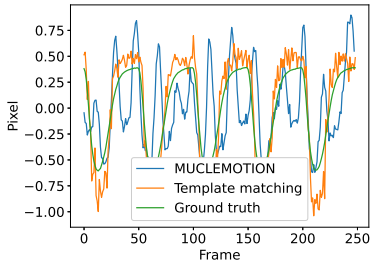

Fig. 60. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 0.5-pixel amplitude and 20% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

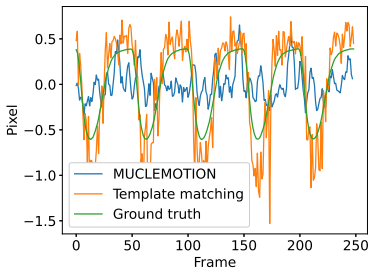

Fig. 61. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 0.5-pixel amplitude and 25% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

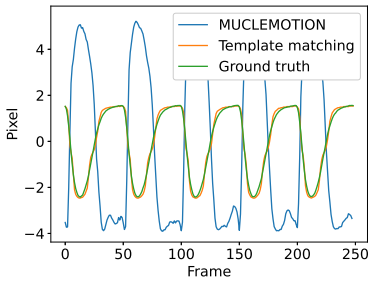

Fig. 62. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 2-pixel amplitude and 1% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

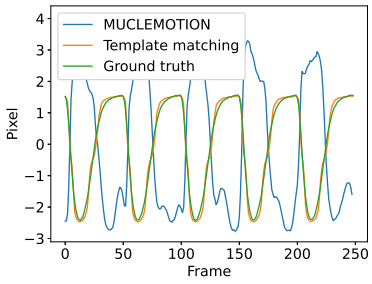

Fig. 63. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 2-pixel amplitude and 5% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

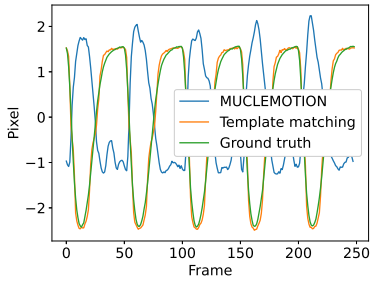

Fig. 64. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 2-pixel amplitude and 10% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

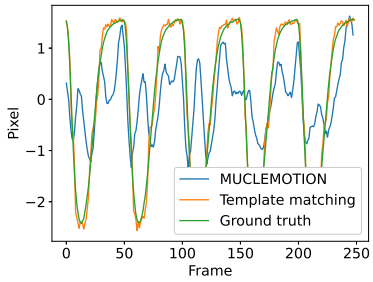

Fig. 65. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 2-pixel amplitude and 20% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.

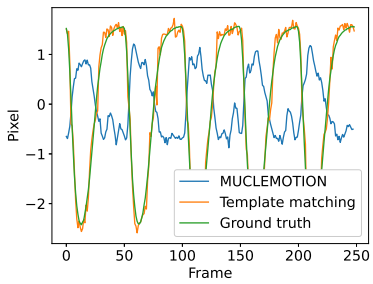

Fig. 66. Comparison of a trajectory from a simulation video of an previously recorded EHT trajectory with a 2-pixel amplitude and 25% noise applied, MUSCLEMOTION versus template matching with sub-pixel precision versus ground truth.