Project report

Article name:

“Arbitrary Order Total Variation for Deformable Image Registration”

By Jinming Duan, Xi Jia, Joseph Bartlett, Wenqi Lu, Zhaowen Qiu

2023, ScienceDirect

Submitters:

Lior Dvir	207334376
Neta Becker	209244839

Contents

Main concept 3
Background 4
Introduction of the given article 4
Object Tracking 5
Method 7
Experimental results 8
Effect of λ 9
Effect of Taylor Series Order 10
Effect of Scale Levels 11
Effect of TV Order 12
Effect of Morphological Operations 14
Results of the compared model 15
Summary 16
References 17

Main concept

The paper we have chosen is: “Arbitrary Order Total Variation for Deformable Image Registration”, by Jinming Duan, Xi Jia, Joseph Bartlett, Wenqi Lu and Zhaowen Qiu.

Image registration is the process of creating alignment between two or more images from various imaging equipment or sensors taken at different times and angles, or from the same scene. The alignment between different images could be rigid and non-rigid. Rigid alignment is an alignment that could be defined with an affine transform, whereas non-rigid alignment could not.

The paper presents a new mathematical model which performs non-rigid image registration, meaning the algorithm can handle alignment between deformed images.

We would like to apply the algorithm and its resulting registration map in the field of object tracking, where objects can change shape and form in time. With an initial mask of an object in a time-series video we can calculate the registration map between consecutive frames and warp the mask through time. Theoretically, that would allow us to get a mask of the tracked object throughout the entire input video even if the object’s form changes between frames.

The goal of this project is to tackle challenging scenarios, such as tracking a tennis ball in a video. See selected frames from tennis matches for example:

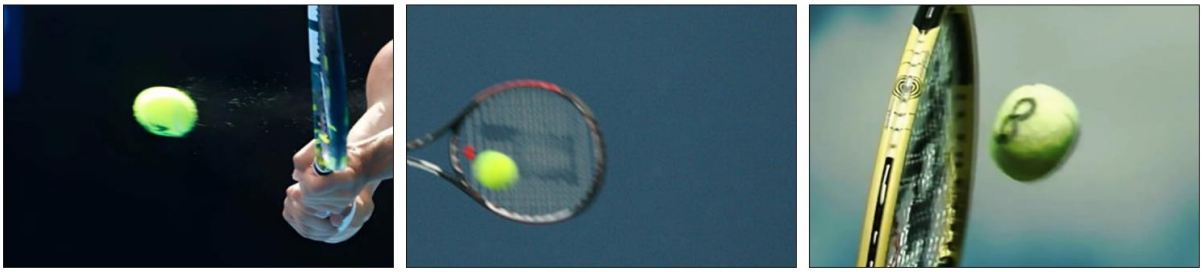

Figure 1

Using a regular fixed mask the shape of a ball might result in “losing” the ball mid-video, since current tracking methods rely on certain assumption; wether it’s the shape of the object, the assumption that the object moves in constant accelaration, the assumption that the scale of the object is constant etc. Using the algorithm presented in the article could act as the solution to that problem.

Background

Introduction of the given article

In the given article, the authors present a mathematical model which performs image registration. That model uses sum of absolute differences (SAD) with an arbitrary order total variation as a regularization term. The Objective is to find u^* that minimizes the following:

$$\min_{u_1,u_2} \int_\Omega \left|I_1\left(x+u(x)\right) - I_0(x)\right| dx + \lambda \int_\Omega \sqrt{\sum_p \left|\nabla^n u_p(x)\right|^2}\, dx$$

Where:

- u is the 2D displacement field
- I_i are the input images
- $\nabla^n u_p$ is the n'th order distributional derivative of u. n is an arbitrary integer.

Note: The authors use the intensity consistency constraint (the intensity at point x in the target image is the same as that at point $x+u$ in the source image) to simplify the model. That assumption is common, specifically in medical imaging, due to two main reasons:

1. In medical imaging we expect matching areas to have the same intensity (e.g., bones in X-ray imaging will appear white).
2. Pre-processing of the input images could reach the wanted state.

Since the optimization of the given term is challenging, the authors propose to approach it in the following matter:

1. Using linearization to simplify the model.
2. Breaking down the optimization process to multiple steps using alternating direction method of multipliers (ADMM)[1].

[1] The alternating direction method of multipliers (ADMM) is an algorithm that attempts to solve a convex optimization problem by breaking it into smaller pieces, each of which will be easier to handle.

The model described in the article was proven to be successful, as can be seen in the following results:

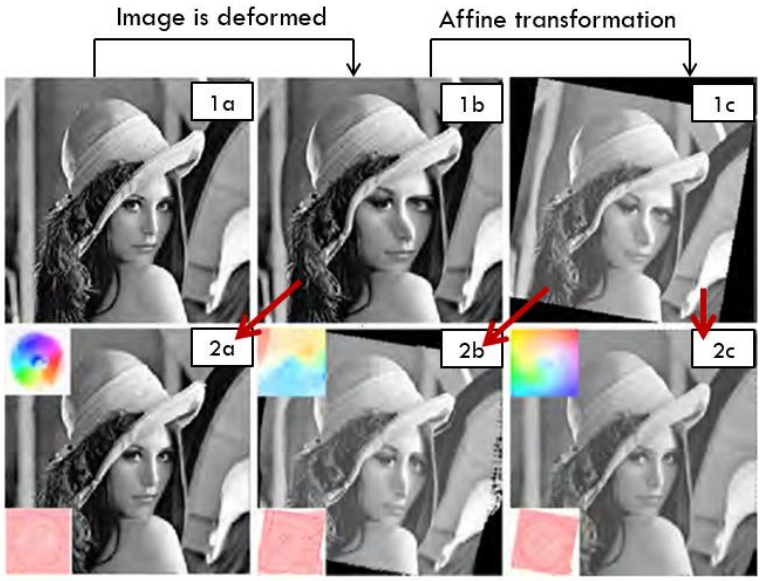

Figure 2

- (1a) Original image
- (1b) The original image after it was deformed
- (1c) The deformed image after an affine transform was applied to it
- (2a) The result of applying a 1st order Total Variation to image (1b)
- (2b) The result of applying 1st order Total Variation to image (1c)
- (2c) The result of applying 3rd order Total Variation to image (1c)

As can be seen in figure 2, the 1st order methods rely on the initial alignment of images, making them better suited when an affine linear pre-registration is possible. However, this is not the case for higher-order methods.

Object Tracking

Object tracking is the process of identifying and tracking objects during a series of frames of a video. The tracking process usually starts with detection of the object of interest, followed by locating the object in the following frames.

Object tracking analyzes each frame in a given video to identify the object of interest and draw a bounding box around it. The object is effectively tracked throughout the video by performing this operation on all frames.

Some of the most known algorithms in the field of object tracking are Kalman filter, KCF (Kernelized Correlation Filters) and YOLOv8 (You Only Look Once).

All of the models mentioned above are all likely to struggle (or even fail) tracking an object that experiences non-constant movements or undergoes deformation of some sort. Since those algorithms rely on constant object acceleration, struggle with changes in the scale of the object and struggle with tracking small objects respectively.

In order to apply object tracking on challenging objects we will have to rely on more sophisticated methods.

Method

To achieve the ability to track objects that may change their shapes in between frames, we are planning to implement the following workflow:

For the first frame:

1. Detect the object of interest using object detection algorithm
2. Create a mask of said object

For any frame that follows:

1. Use the algorithm from the given article to shape the mask using the current frame and the previous one as input images
2. Create a mask according to the result of the algorithm
3. Apply “Open” and “Close” morphological operations to fix mask anomalies

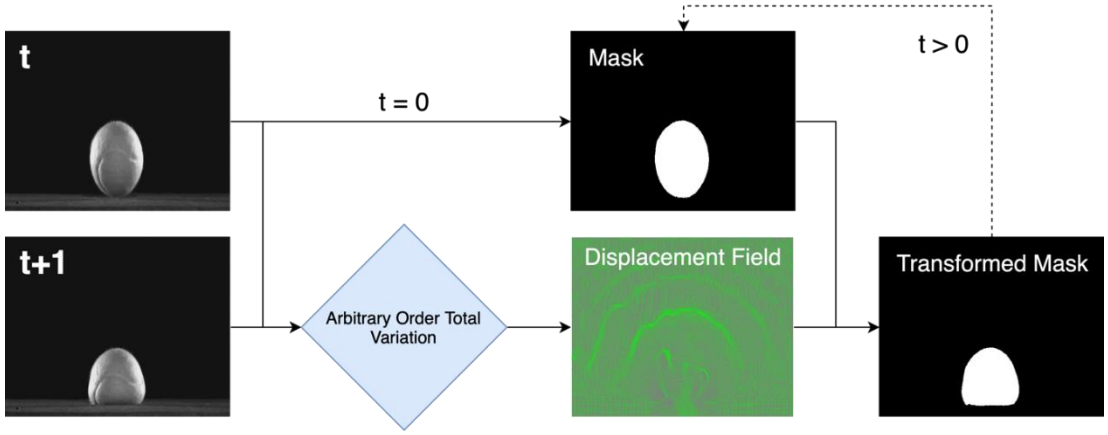

Figure 3

At the first-time step of the algorithm, we extract the target’s mask (A binary image of the same size as the target’s image – with ones where the target is and zeros otherwise) from the first image. This can be done with an object detection algorithm coupled with segmentation, for instance. In this project we focus on object tracking, rather than detecting, so for our purposes we’ve extracted the mask manually.

Each time step *t* we use the Arbitrary Order Total Variation algorithm to calculate the displacement field between image *t* and image *t+1*, which by applying its per-pixel displacement on image *t* will transform image *t* into image *t+1*.

Then, we apply the displacement field on the target’s mask from image *t*, transforming it into the mask of the same object in image *t+1*.

Finally, morphological operations are applied to the transformed mask to address stark deformities caused by the displacement field.

Experimental results

The algorithm includes various parameters that can be tuned:

λ – Affects the rigidity of the displacement field. The higher λ is, the displacement field is more restrained.

Taylor Series Order – The data term is linearized with a Taylor series to make the model convex. The higher the order the, the more accurate the series is to the original data term.

Scale Levels – The algorithm works on different scale levels to accelerate convergence and utilize course information correlation (focusing on general high frequency details such as shapes as opposed to fine details such as texture).

TV Order – The power order over the absolute of image derivatives. First order is ordinary TV. Higher order derivatives allow for smooth-inducing and discontinuities regularizations as well as being less dependent on initial alignment between images.

The following graphs explore the impact of each parameter over the resulting IOU of a short 10-frame video of a tennis ball falling at high speed.

Below are the frames tested on. The challenge posed by the video is misalignment of the target between consecutive frames as well as deformities in its shape.

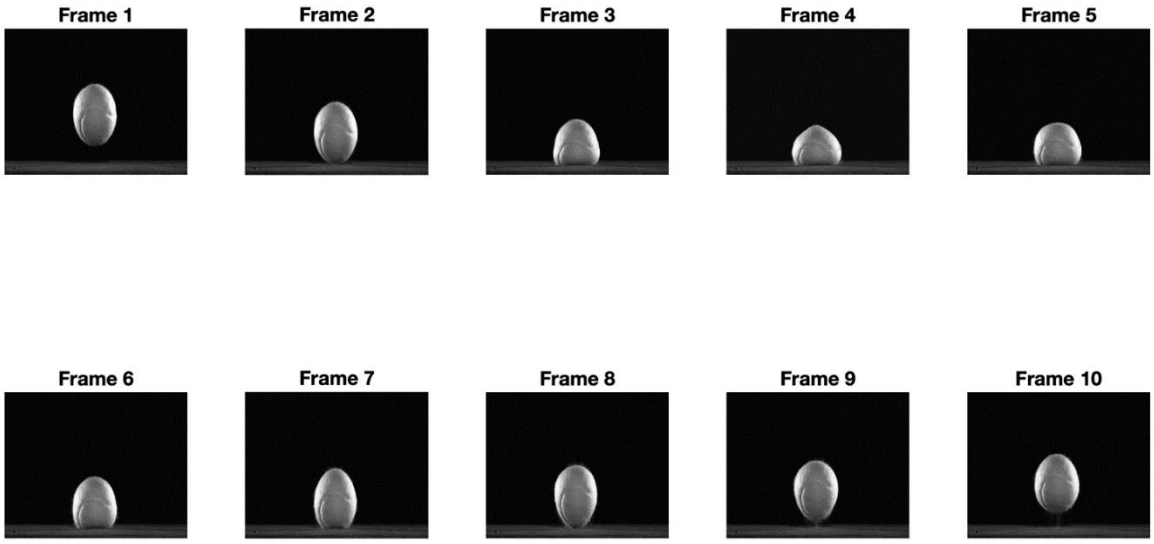

Figure 4

Effect of λ

We've tested different values for λ = {1, 10, 100, 1000, 10000}

Other fixed values:

- Taylor Series Order = 10
- Scale Levels = [16, 8, 4, 2, 1]
- TV Order = 2nd order

The following is the IOU between the algorithm's mask and ground truth mask in the last image for different λ:

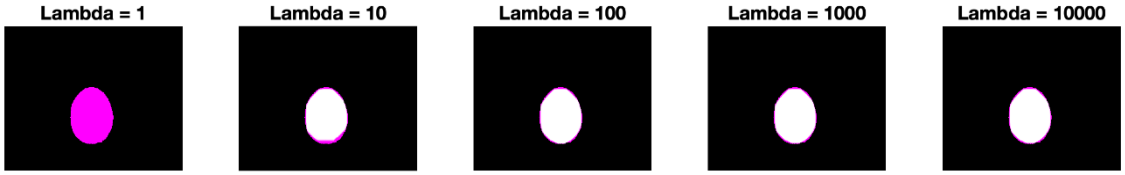

Figure 5

(White is overlap. Magenta is GT mask. Green is algorithm mask)

The following is a graph of IOU value as a function of λ:

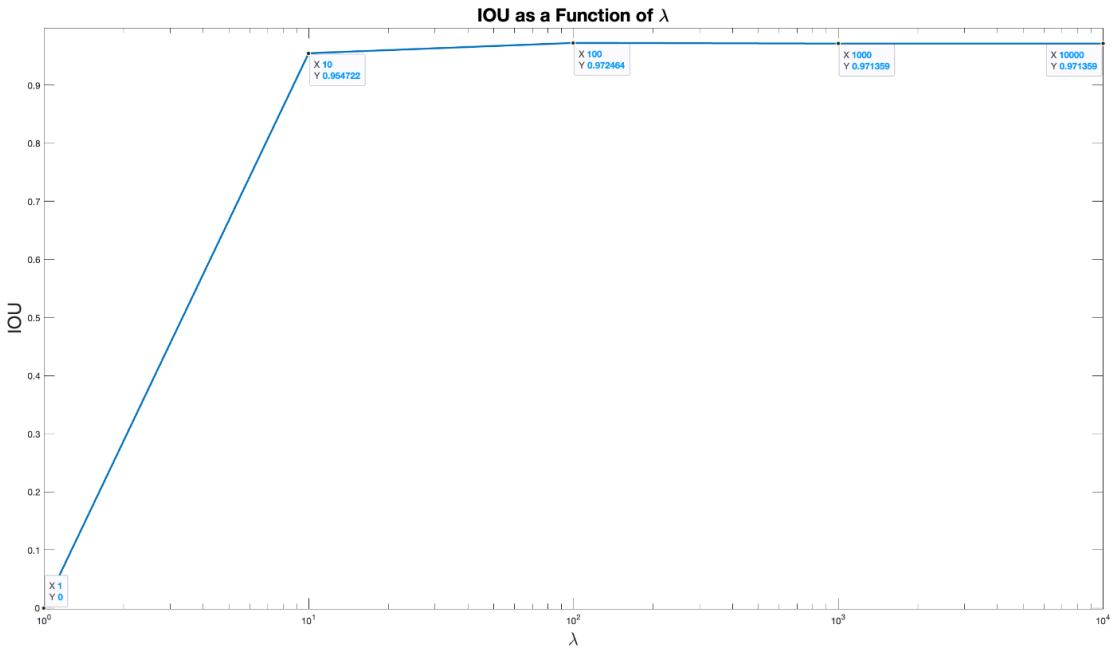

Figure 6

After $\lambda = 10$ we reach a plateau, however the algorithm takes longer to converge the higher λ is. Therefore, for later tests the value $\lambda = 10$ is taken.

Effect of Taylor Series Order

We've tested different Taylor Series Order values = {1, 3, 5, 7, 10, 15}

Other fixed values:

- $\lambda = 10$
- Scale Levels = [16, 8, 4, 2, 1]
- TV Order = 2^{nd} order

The following is the IOU between the algorithm's mask and ground truth mask in the last image for different Taylor Series Orders:

Figure 7

In the algorithm, when convergence of the data term is checked, the Taylor approximation is used. The higher the order the more precise the convergence is. We can see that the resulting masks are more accurate the higher the order, but with diminishing returns as opposed to the time of convergence.

The following is a graph of IOU value as a function of the Taylor Series order:

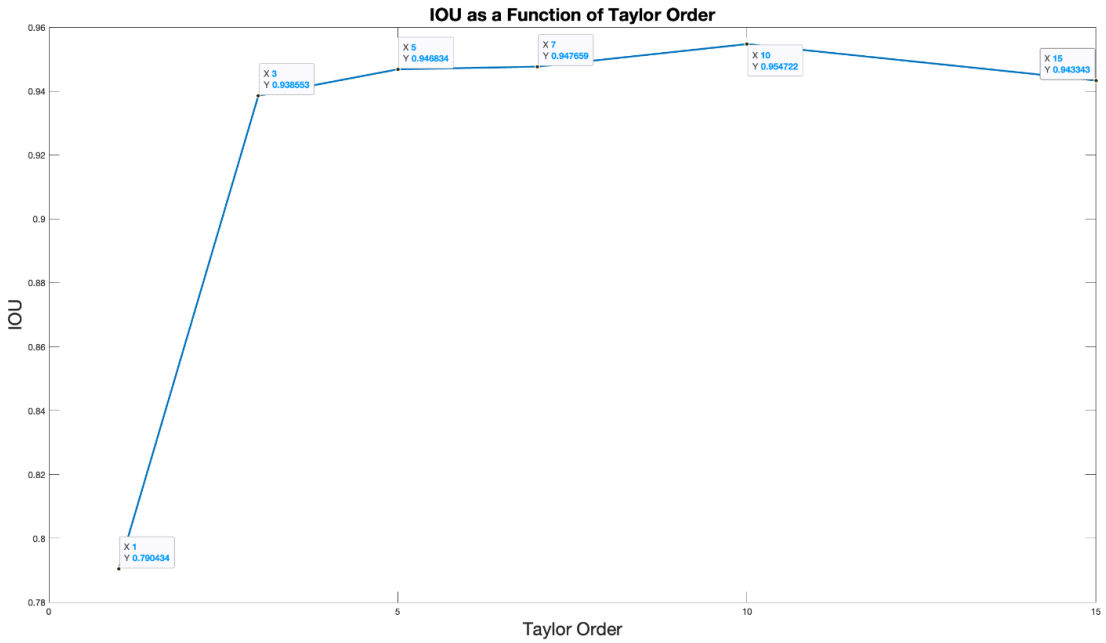

Figure 8

Best results were received for Taylor Series Order = 10, and that is the value we fixed for the rest of the tests.

Effect of Scale Levels

We've tested different pyramid maximum scale levels= {1, 2, 4, 8, 16}

<u>Other fixed values:</u>

- $\lambda = 10$
- Taylor Series Order = 10
- TV Order = 2^{nd} order

The following is the IOU between the algorithm's mask and ground truth mask in the last image for different sets of scale levels (each one starts from the highest one and down samples by a factor of 2 until reaching 1):

Figure 9

And the IOU values themselves as a function of maximum scale level:

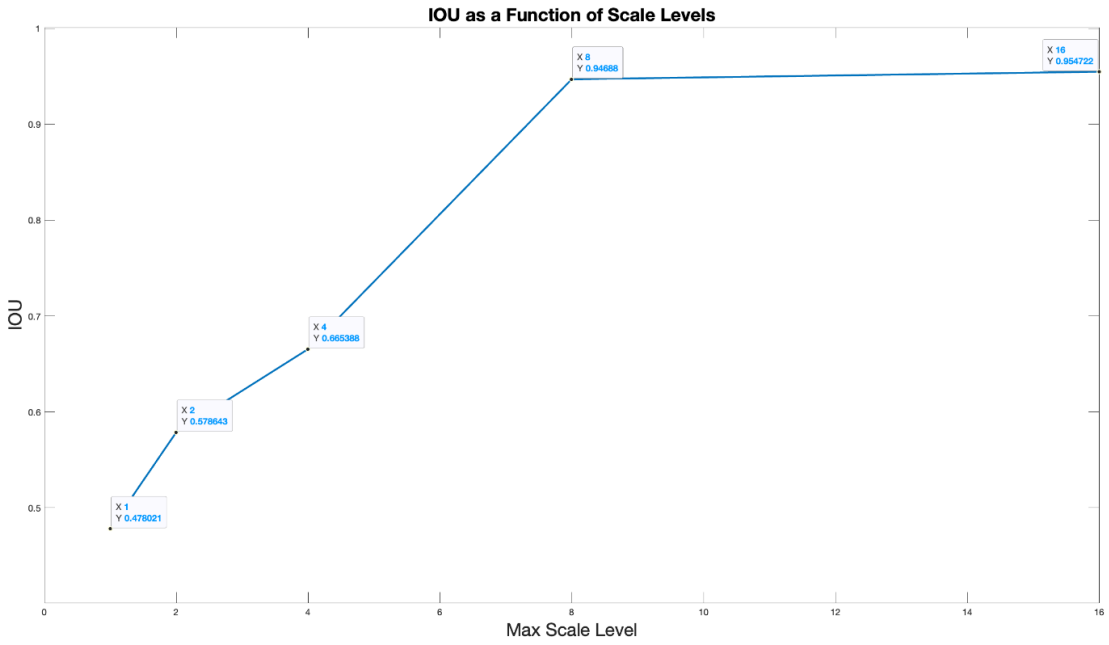

Figure 10

Down-sampling the images with higher scales levels helps the iterative process by first matching the images using their "low frequency" features, such as shape and general pixel values.
Once this is done, the images are better aligned, which in turn leaves the rest of the process focus on the finer details such as texture.

As mentioned before, initial alignment is important to the success of TV flow methods.

Effect of TV Order

We've tested different TV orders= {1st Order, 2nd Order, 3rd Order, 4th Order}

Other fixed values:

- λ = 10
- Taylor Series Order = 10
- Scale Levels = [16, 8, 4, 2, 1]

The following is the IOU between the algorithm's mask and ground truth mask in the last image for different TV orders:

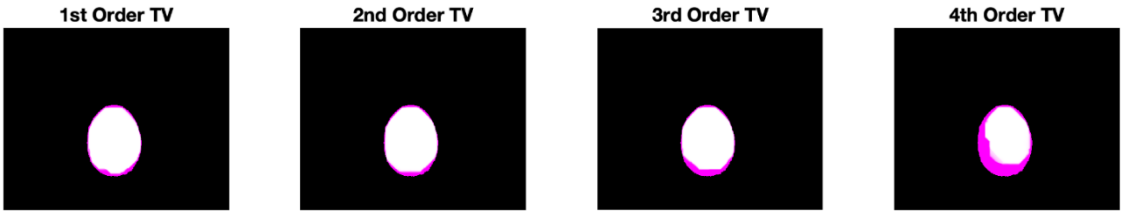

Figure 11

The higher the order the more robust the algorithm is to misalignment of images and its ability to preserve discontinuities, while inducing smoothness. However, as the order gets too high it appears it is too sensitive to discontinuities and produces wrong masks.

The following is a graph of IOU value as a function of TV order:

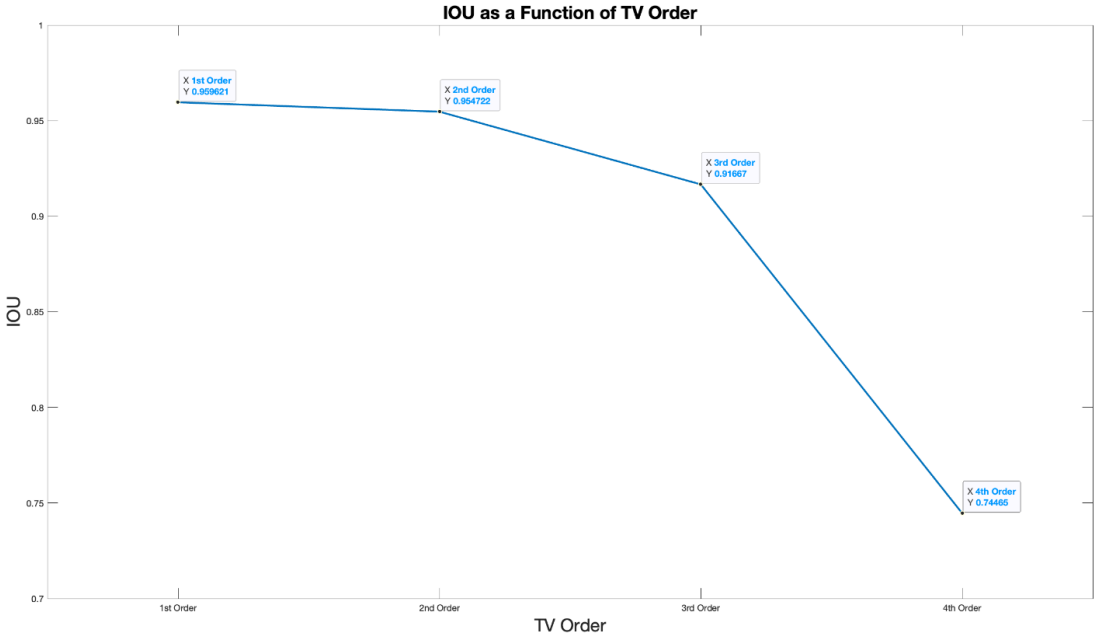

Figure 12

It also seems as if the 1st order TV – Which is regular TV term as learned in class – produced a good mask. However, when a simpler video is used, one where there is only translation between the frames with no change to form:

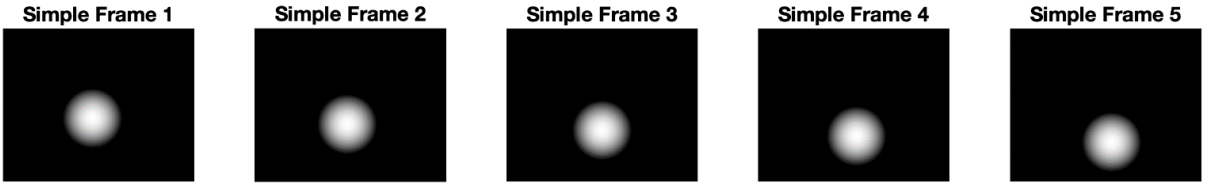

Figure 13

We can see that 1st order TV is left lacking:

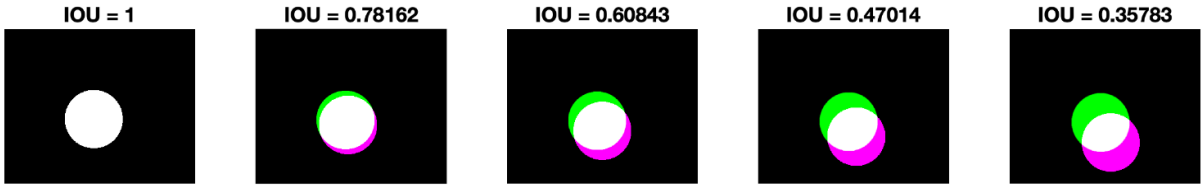

Figure 14

While 2nd order TV handles the translation better:

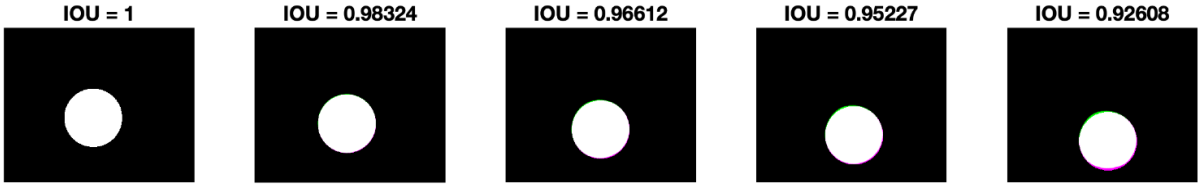

Figure 15

The data term alone is not enough to reach correct translation through gradient descent. 1st order TV is also not enough as a regularization term.
As mentioned by the paper, TV suffers when the frames misalign by an affine translation. This is such a case where using higher order TV, which handles the translation successfully, showcases this point.

Effect of Morphological Operations

In the algorithm, after the mask is transformed via the displacement field, we apply a morphological operation of **Open** on it (An **Erosion** followed by **Dilation**). Its effect is removing small protrusions and deformities in the mask. We also note that the morphological operation of **Close** might have been useful if small holes in the mask were to appear, which they haven't.

The following set of images show the results of the algorithm using regularization of 1st order TV with no morphological operations:

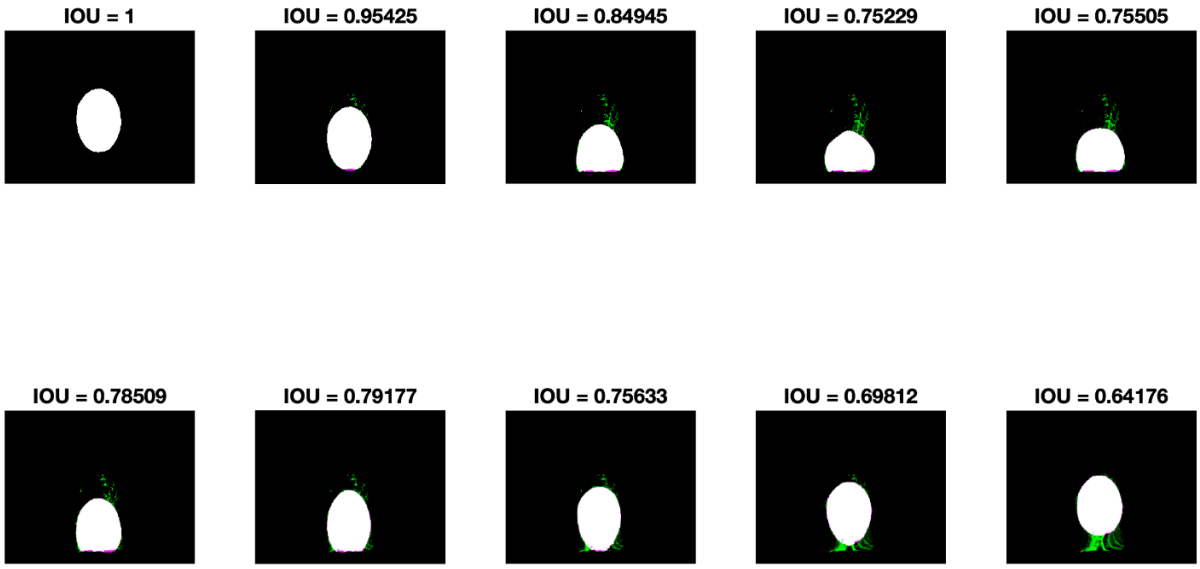

Figure 16

As opposed to 1st order with Open operation:

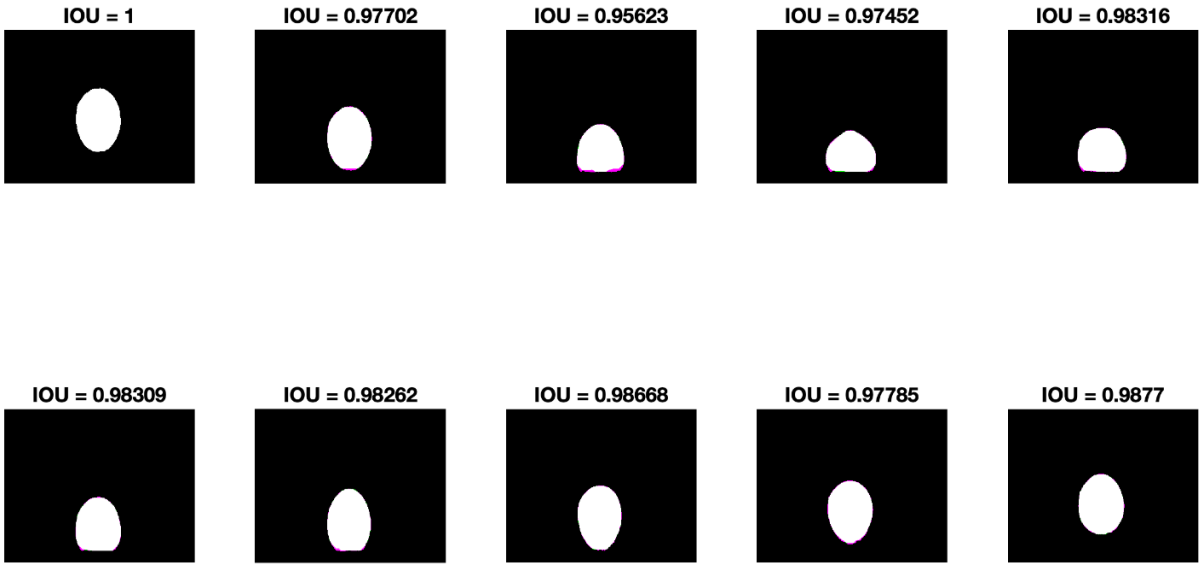

Figure 17

TV as a regularization term benefits the algorithm by making it robust to change in form. However, that same sensitivity can be detrimental, such as in the Tennis ball example – where dust and hairs mix and flow together ending with hairline deformities in the mask. These the morphological operation can easily dispatch.

Results of the compared model

The compared model was written by Adrian Rosebrock. It's an object tracking algorithm that was developed specifically for tracking tennis balls. The algorithm relies on the shape and color of a tennis ball (meaning if the input video is in grey scale – the algorithm will fail to detect the initial object).

The results of the algorithm are:

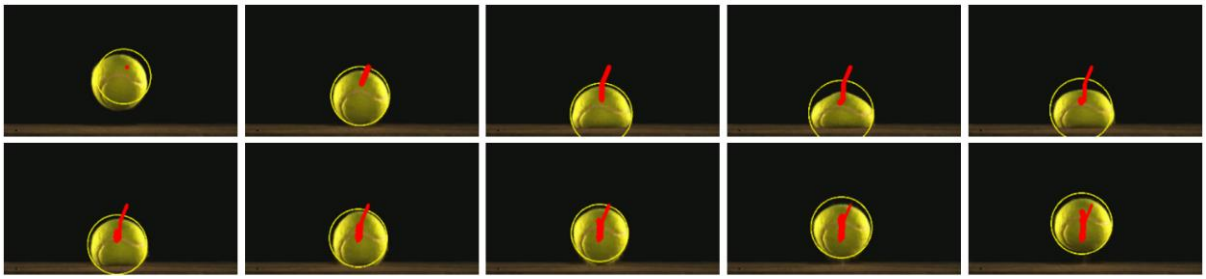

Figure 18

(The red line represents the movement of the object).

Though it succeeds on the given video and runs much faster than our proposed algorithm, the presented algorithm isnt as versatile as the algorithm we propose. This algorithm relies on a circle shaped mask to track the object, that is achieved by searching the shape and color of a tennis ball in the frames. Therefore even though it achieves good results, its not as robust as the algorithm we propose.

Summary

The algorithm, based on the Arbitrary TV method from the paper, successfully functions as an object tracker.

It manages to handle an object's form changing over time and fix deformities in its mask.

However, there is no correct-works-for-every-scenario set of parameters. Videos with different characteristics need different TV orders, λ values, etc.

The algorithm is also very slow. Several minutes until convergence between frames. Which makes it non-feasible as a real-time tracker.

The field of tracking algorithms is already very competitive with the superior technology of neural networks which do work in real-time.

Therefore, in our opinion, the strong competition together with its faults makes our algorithm an interesting prospective, but not one that can reach SOTA status and tour conferences.

References

1. ‘’Arbitrary Order Total Variation for Deformable Image Registration’’
   Jinming Duan, Xi Jia, Joseph Bartlett, Wenqi Lu and Zhaowen Qiu, 2023, ScienceDirect
2. “Ball Tracking with OpenCV”
   Adrian Rosebrock, 2015
3. “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers”
   Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, 2010, now (the essence of knowledge)