Computer Vision Final Course Project: The Auto-Bio Challenge

Jiarui Li*
College of Engineering
Peking University
lijiarui@pku.edu.cn
Github Repository: github.com/LiJiarui111/Trans2Seg-AutoBio-Challenge

Abstract

Object segmentation and object tracking are quintessential tasks in the computer vision field, but most of the work nowadays focuses on only the opaque object rather than those transparent ones, even though there are many scenarios transparent objects are pervasively used, such as in the biology laboratory. In this project, we aim to propound a biological experiment monitoring system. To reach this destination, we conclude and use the current SOTA segmentation and object tracking model to deal with the formulated AutoBio Challenge, which consists of transparent object segmentation and liquid tracking (transparent object tracking). We test the baseline model and modify the network pipeline and structure from the papers to make our model perform better in our environment. The results show that our model can successfully deal with the segmentation problem and can pave the way to make it possible for the biological experimental monitoring system.

1 Introduction

As an experimental discipline, biology and those biological experiments are notoriously time-consuming and error-prone, so a proper vision-based monitoring system is essential. However, due to the variety of equipment used in biological experiments and much of the equipment being transparent and tiny in scale, this brings great difficulties to the monitoring system as most of the current vision and understanding models may find it challenging to deal with transparent objects because the transparency nature that may mix the objects with their surrounding background, and confuse the models in segmentation and other aspects. In this project, we explored and tried the baseline segmentation method, fork and modified a novel transformer-based pipeline, successfully dealt with our AutoBio experimental scene, and segmented the transparent experimental equipment out of the background. Further explorations about the liquid tracking task are also carried out, which paves the way for fully solving this challenge.

1.1 Problem

The AutoBio Challenge aims to help biology experimenters to detect errors in experiments and help the experimenters in improving their success rate. To reach this destination, first, we should segment the transparent objects from the background, which are pretty pervasive in the biology lab. Secondly, liquid tracking is a common experimental behavior and is very error-prone, so correctly understanding the transition of liquid is the basis for effective error detection. Moreover, the understanding of cooperation is important when dealing with some long-term and complex experiments, so our model should have the ability to understand the experiment process and the cooperation relationship to assist the experimenter in avoiding making mistakes.

*Website: https://jrli.org/,
Pretrained model: https://disk.pku.edu.cn:443/link/309A59AC8A6C5AEFB7EB46D8815C82A8,
Code: https://github.com/LiJiarui111/Trans2Seg-AutoBio-Challenge

(a) **Glass cup**. The transparent cup is put in front of a stone wall.

(b) **Transparent plastic bottle**. Those black clots on the bottle are opaque stickers.

Figure 1: **The results of the Trans2Seg model.** After training 150 epochs on the *Trans10k_v2* dataset, our model can successfully segment transparent objects from its background.

Table 1: **Baseline training results.** The accuracy and mIoU of the evaluation set and the test set.

Mode	Accuracy	mIoU
Evaluation	0.92512	66.709
Test	0.92166	60.760

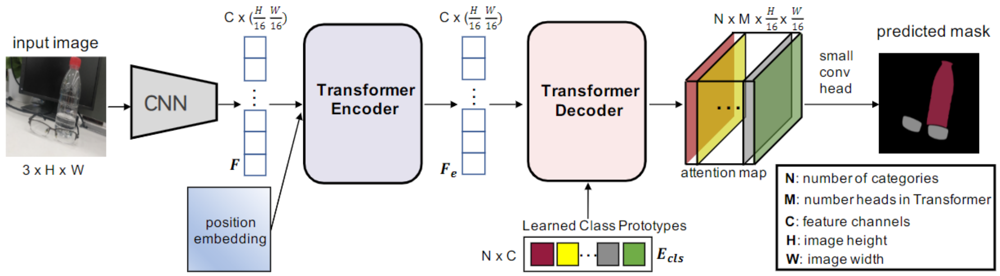
1.2 Related Work

This Auto-Bio Challenge comprises two main parts, transparent objects segmentation and experiential video understanding (liquid tracking and cooperation understanding).

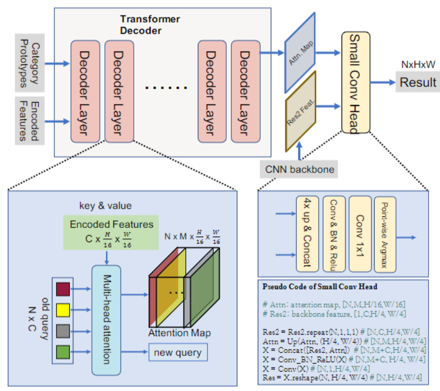
Transparent Object Segmentation. Object segmentation has been a crucial computer vision task for a long time. There are many kinds of segmentation. The difference between those types lies in sophistication. Semantic segmentation classifies pixels according to the category of the object they belong to and has attracted lots of attention in recent years. The deep-learning-based method FCN [7] came up with a fully-convolution architecture and took semantic segmentation to a new stage. After that, many augmented methods have been proposed, such as xxx, xxx and xxx.

Transparent object segmentation is unique since many effective segmentation algorithms may fail in this task because the transparent objects' transparent nature, fusion with background, and flexible reflection, refraction, and light projection may confuse the sensors. Some methods based on optical sensor information are proposed, such as [10, 16, 17, 5]. Those methods used optical sensor information like polarization cues to compensate for the drawback and had good results. However, Pure RGB-based transparent object segmentation still remains to be a problem. The lack of transparent object datasets is also a crucial problem. In 2020 and 2021, two efficient datasets *Trans10K* [13] and *Trans10k_v2* [15] were proposed. Myriad works concentrate on transparent object segmentation, and some works focus on CNN-based complex network structure to detect and integrate contextual features to handle the transparency difficulties [1, 9], but those kinds of methods have strict conditions of use, for example, need complete transparent glass or can only segment glass door or wall, and cannot be applied to real-world scenarios. Recently many Transformer-based [11] algorithms have been proposed to address the transparent object segmentation problem [14, 15, 18] and showed considerable effects. The attention mechanism can effectively augment the model's ability in difficult object segmentation problems and is quite promising.

Transparent Object Tracking. Object tracking is an important problem in computer vision, closely related to many other fields, such as robotics. There are many existing works focused on opaque object tracking and had satisfactory results, and in [12], a popular Object Tracking Benchmark (OTB) is proposed. However, when it comes to transparent objects, taking liquid, for example, things will become quite different. Similar to object segmentation, the transparent feature also causes a lot of trouble to existing algorithms. In fact, transparent object tracking is still a mainly untouched area compared with opaque object tracking. In 2021, [4] proposed a Trasparent Objects Tracking Benchmark (TOTB) and implemented a simple but effective transparent object tracking framework TransATOM. The TransATOM model uses the ATOM architecture [2] together with the transparent features extracted by the transparent object segmentation model, and here we use the model in [13, 15] and integrate common features and special transparent features. There is another interesting argument



(a) **Whole pipeline**



(b) **Details of the encoder and decoder**

Figure 2: **The whole pipeline of the hybrid CNN-Transformer architecture in [15]** and the details of transformer encoder and decoder. These figures are copied from [15]. This pipeline used a CNN backbone and transformer encoders and decoders to extract effective features and self-attention. Used a small convolution head to predict the category probability distribution of each pixel.

about the result of the TOTB paper discussed above. According to [8], actually, in the TOTB paper, the comparisons have not taken the training set into consideration. If we adjust the training set and concentrate more on the transparent objects, the current State-Of-The-Art model can have better results than the ones shown in the TOTB paper's comparing section. Understanding is a popular area in computer vision or artificial intelligence.

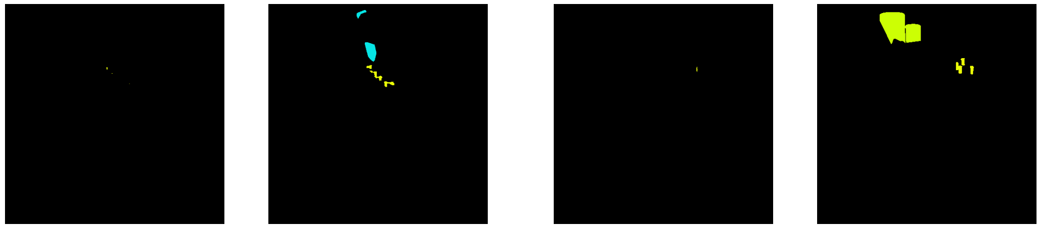
1.3 Overview

We organize the remainder of the report as follows. Sec. 2 describes our transparent object segmentation method and the results on the biology lab dataset. Sec. 3 describe the method we use to deal with liquid tracking problem. Finally, we conclude the paper in Sec. 4.

2 Transparent Object Segmentation

Faced with this problem, we got the first idea to reproduce a baseline and observe the result on our experimental dataset. At first, we select some popular object segmentation pipelines and apply those baselines to our problem, but the output is very dissatisfying. The model cannot segment the devices, especially those tiny-transparent objects, as shown in Fig. 3. It can be deduced that transparency is a critical problem. To deal with this problem, we chose a unique transparent object segmentation network, the **Trans2Seg** model in [15]. Although there are other models, such as [18], which had better results (accuracy, mIoU) than Trans2Seg, however (i) Those new models didn't improve much, and (ii) Trans2Seg's code is open source, and the code's structure is pretty straightforward, which can make my reproduce and debug process much easier and save a lot of time. We cloned the GitHub project from the paper's official project website [2], but it seems that the code has some bugs because we confronted with some errors and there are many unclosed and unanswered issues which depict

[2]https://github.com/xieenze/Trans2Seg



(a) *our baseline result I* (b) *ground truth I* (d) *our baseline result II* (e) *ground truth II*

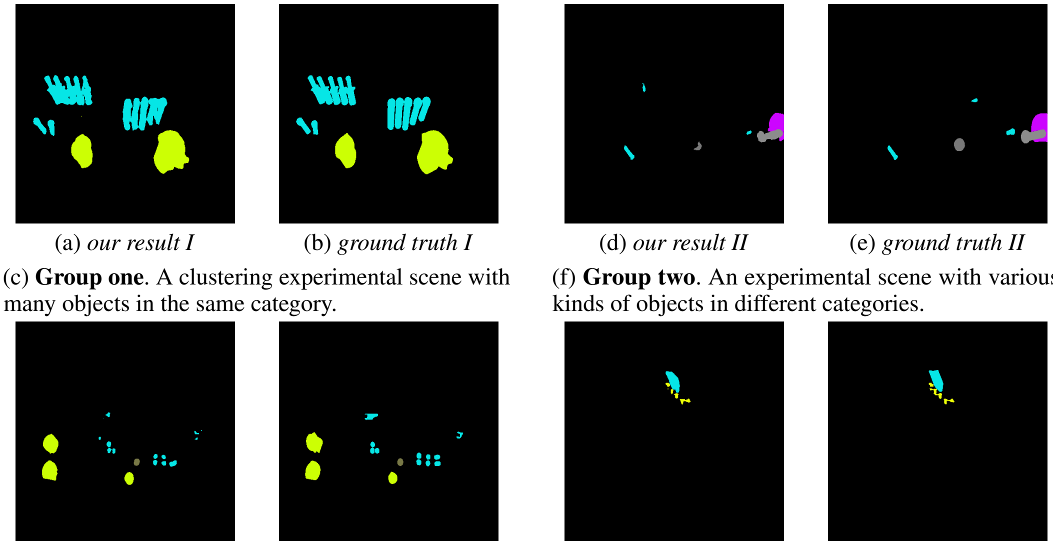
(c) **Group one** (f) **Group two**

Figure 3: **The results of the original baseline model.** Even after tuning and modifying, the baseline classical segmentation network cannot successfully deal with our experimental scene.

similar problems on Github. There are so many bugs that we spent several days struggling in this sea of bugs, which is extremely painful. After an arduous struggle, we finally solved the problems and made the training process start to run. My original plan is to train the model on *Trans10k_v2* dataset, fix the weight of some layers of the trained model, and re-train the model on our experimental dataset. As we know, training a transformer model requires a large dataset [3], but our experimental dataset is obviously too small for the transformer's training. By training on a larger dataset in advance, we can equip our model with better feature extraction abilities and can be expected to have a better result on our dataset.

We trained the model on *Trans10k_v2* for 50 epochs. However, we use this model to make inferences and visualize the outputs. Obviously, the outputs are very dissatisfying. So we extracted the training process to 150 epochs and tried again. After several days of struggle, the result is shown in Tab. 1, and the demo figures are shown in Fig. 1. Besides, we forked the original GitHub project and uploaded my bug-fixed and visualizing-enabled version to facilitate later users.

The next step is to train the Trans2Seg model on our AutoBio Challenge dataset. Unlike the dataset format in Trans2Seg's original project, our dataset is organized in COCO format [6], so the first step is to utilize the dataloader code and apply our dataset to the training and testing pipeline. We modified the network structure and tried the training from scratch, but the results showed that it seems that some tiny objects were still missed, even though the results were already much better than the baseline model showed in Fig. 3, and some categories had much worse results than others. For the latter problem, the reason is that not all the experiment equipment (entities) are displayed in our dataset, or some just exist in the test set, so our model cannot learn knowledge of those object categories. For the previous problem, we tried various ways to solve it, which we will elucidate as follows. First, considering our dataset may be too small, we may have to use data augment techniques to guarantee that the model can extract useful transparent-related features. As there are so many tiny objects in our scene, data augmentation became more critical. Also, we tried to modify the structure of the pipeline shown in Sec. 1.2. The things we have done can be divided into four categories. The first thing to do is to change the number of output dimensions. In the original paper, we have 12 categories, but here we have 21 categories (both include the background). Secondly, we tried to change the backbone ResNet category from ResNet50 to ResNet101, but the result's improvement is subtle and unrepeatable. We also tried other modifications, which changed the modes of some modules in the pipeline, but no noteworthy difference can be made. It is reasonable for us to assert that to enhance the model's performance further, some significant modifications to the structure of the network (the pipeline) are necessary. Thirdly, we tried to change the CNN feature extract module into a pure transformer module, but the results were dissatisfying, and it seems that the code still had some bugs unsolved. We struggled for several days and finally moved back to the original CNN extractor. Fourthly, as the AutoBio Challenge requires us to make a monitoring system of the experimental scene, only concentrating on those experiment facilities may not be enough. Other objects such as tables and shelves, even gloves and alcohol blowtorches. The current segmentation pipeline was designed for segmenting transparent objects, so we can add the ability to make it adapt to those opaque objects. Here we designed a double-head segmentation framework. Based on the pipeline shown in Sec. 1.2, we added another successful classical object segmentation model parallel to the original one and decided the affiliation of each pixel at the last step. We use pure probability comparison to judge the affiliation of the pixels, but the results showed that the classical way always got the higher probability, so it is actually a conundrum to mix both results, and we are still trying to enhance it.



(a) *our result I* (b) *ground truth I*

(c) **Group one**. A clustering experimental scene with many objects in the same category.

(d) *our result II* (e) *ground truth II*

(f) **Group two**. An experimental scene with various kinds of objects in different categories.

(g) *our result III* (h) *ground truth III*

(i) **Group three**. A quintessential scene with many dot-like tiny objects.

(j) *our result IV* (k) *ground truth IV*

(l) **Group four**. A scene that only has few objects lie in a small part of the scene.

Figure 4: **The results of our AutoBio transparent object segmentation model.** After training 150 epochs on the *Trans10k_v2* and on our experimental scene dataset, our model can successfully segment transparent objects from its background.

Table 2: **Autobio transparent object segmentation training results.** The category names and IoUs are listed below. Here we omit all the zeros as not all the categories have object entities in our training set.

Class name	IoU
Background	0.993482
PCR Tube	0.387770
Tube	0.547398
Waste box	0.648489
Vial	0.464103
Measuring flask	0.604211
Water bottle	0.879246
Erlenmeyer flask	0.522936

In training, we used the Trans2Seg model trained by *Trans10k_v2* dataset, froze the first few layers, and retrained other layers on our dataset. After 100 epochs of training, the results are shown in Fig. 4. According to the results figures, our model could efficiently segment the objects we need out of the model in most of the cases, but there is still one problem. In our training and testing data, some object category has few figures, and the scale of data is not enough for training, especially for our transformer-based model.

3 Liquid Tracking

First, we tried reproducing the baseline of TransATOM mentioned in the TOTB paper [4]. The transparent object segmentation model is an essential part of the tracking model. According to the baseline, we use the baseline model of *Trans10k* dataset in Sec. 2 for the first trial. After implementing the tracking model, the baseline testing results are shown in **??**. The results show that the method mentioned in the TOTB paper is actually very effective. The object tracking framework is based on the pytracker repository

Considering our task, we aim to analyze a given video from its frames, identify the liquid in the containers and get a "trajectory" of a particular liquid. We slightly modify the pipeline and apply the model to our dataset. The salient point is that the liquid could be bifurcate in our task, which means that the experimenter may spill the liquid and pour the liquid into different tubes or bottles. There are too many bugs in the code, and we are still struggling and analyzing.

4 Conclusion

In this project, We explored the AutoBio Challenge, which aims to set up a vision-based experimental monitoring system to prevent errors in the experiment process. Specially, we tried to solve the transparent object segmentation, and liquid tracking (transparent object tracking) problems, which are quite different from classical ones, for transparency is pretty challenging to deal with. We reproduced and tested the baseline model. After that, we fork and modified the network pipeline from paper, tried various ways, and finally made our model perform well in our AutoBio task in comparison to the baseline results. Our results show that our model can successfully segment those tiny transparent objects from the background. We further explored and implemented the transparent object tracking pipeline and paved the way for fully solving the AutoBio Challenge.

In the final paragraph of this report, the author would like to apologize that all the "we" in this report actually refer to only one person, that is me, as my team only have one member, and the use of "I" in the academic report seems to be informal. I will give you my most sincere apologies if this confuses you during your reading.

References

[1] Guanying Chen, Kai Han, and Kwan-Yee K Wong. Tom-net: Learning transparent object matting from a single image. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9233–9241, 2018. 2

[2] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. Atom: Accurate tracking by overlap maximization. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4660–4669, 2019. 2

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 4

[4] Heng Fan, Halady Akhilesha Miththanthaya, Siranjiv Ramana Rajan, Xiaoqiong Liu, Zhilin Zou, Yuewei Lin, Haibin Ling, et al. Transparent object tracking benchmark. In *International Conference on Computer Vision (ICCV)*, pages 10734–10743, 2021. 2, 5

[5] Agastya Kalra, Vage Taamazyan, Supreeth Krishna Rao, Kartik Venkataraman, Ramesh Raskar, and Achuta Kadambi. Deep polarization cues for transparent object segmentation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8602–8611, 2020. 2

[6] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision (ECCV)*, pages 740–755. Springer, 2014. 4

[7] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015. 2

[8] Alan Lukezic, Ziga Trojer, Jiri Matas, and Matej Kristan. Trans2k: Unlocking the power of deep models for transparent object tracking. In *British Machine Vision Conference (BMVC)*, 2022. 3

[9] Haiyang Mei, Xin Yang, Yang Wang, Yuanyuan Liu, Shengfeng He, Qiang Zhang, Xiaopeng Wei, and Rynson WH Lau. Don't hit me! glass detection in real-world scenes. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3687–3696, 2020. 2

[10] Haiyang Mei, Bo Dong, Wen Dong, Jiaxi Yang, Seung-Hwan Baek, Felix Heide, Pieter Peers, Xiaopeng Wei, and Xin Yang. Glass segmentation using intensity and spectral polarization cues. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12622–12631, 2022. 2

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017. 2

[12] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2411–2418, 2013. 2

[13] Enze Xie, Wenjia Wang, Wenhai Wang, Mingyu Ding, Chunhua Shen, and Ping Luo. Segmenting transparent objects in the wild. In *European Conference on Computer Vision (ECCV)*, pages 696–711. Springer, 2020. 2

[14] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:12077–12090, 2021. 2

[15] Enze Xie, Wenjia Wang, Wenhai Wang, Peize Sun, Hang Xu, Ding Liang, and Ping Luo. Segmenting transparent object in the wild with transformer. *arXiv preprint arXiv:2101.08461*, 2021. 2, 3

[16] Yichao Xu, Hajime Nagahara, Atsushi Shimada, and Rin-ichiro Taniguchi. Transcut: Transparent object segmentation from a light-field image. In *International Conference on Computer Vision (ICCV)*, pages 3442–3450, 2015. 2

[17] Yichao Xu, Hajime Nagahara, Atsushi Shimada, and Rin-ichiro Taniguchi. Transcut2: Transparent object segmentation from a light-field image. *IEEE Transactions on Computational Imaging*, 5(3):465–477, 2019. 2

[18] Jiaming Zhang, Kailun Yang, Angela Constantinescu, Kunyu Peng, Karin Müller, and Rainer Stiefelhagen. Trans4trans: Efficient transformer for transparent object segmentation to help visually impaired people navigate in the real world. In *International Conference on Computer Vision (ICCV)*, pages 1760–1770, 2021. 2, 3