SRS_CIS-419-401 2022A Final Project Report

Zhenglin Zhang, Yunqi Xu, Yuxin Wang

TOTAL POINTS

14 / 15

QUESTION 1

1 Grading [select all pages for this question]

14 / 15

+ **0 pts** **Q1. Does the final report match the provided template?** No

✓ **+ 1 pts** **Q1. Does the final report match the provided template?** Yes

+ **0 pts** **Q2. To what extent has feedback from previous rounds been effectively addressed?** e.g. Ignored Completely

+ **1 pts** **Q2. To what extent has feedback from previous rounds been effectively addressed?** e.g. barely addressed anything substantial

+ **2 pts** **Q2. To what extent has feedback from previous rounds been effectively addressed?**e.g. clear response to feedback, but ignored some key feedback

✓ **+ 3 pts** **Q2. To what extent has feedback from previous rounds been effectively addressed?** e.g. addressed all key feedback from both previous rounds

+ **0 pts** **Q3. To what extent does this report identify and target a novel contribution, which it sufficiently differentiates from prior work** e.g. clearly no novel contribution at all,

+ **1 pts** **Q3. To what extent does this report identify and target a novel contribution, which it sufficiently differentiates from prior work**e.g. some novelty likely, but no clear differentiation within the report.

✓ **+ 2 pts** **Q3. To what extent does this report identify and target a novel contribution, which it sufficiently differentiates from prior work** e.g. clearly novel contribution, well differentiated from prior work

+ **0 pts** **Q4: Did the team execute a contribution-focused, step-by-step work plan, sharing workload among team members, mitigating obvious risks, as needed?** e.g. no clear plan at all, or barely any execution.

+ **1 pts** **Q4: Did the team execute a contribution-focused, step-by-step work plan, sharing workload among team members, mitigating obvious risks, as needed?** e.g. good plan, but limited progress towards execution

✓ **+ 2 pts** **Q4: Did the team execute a contribution-focused, step-by-step work plan, sharing workload among team members, mitigating obvious risks, as needed?** e.g. good plan, and executed well and almost to completion

+ **0 pts** **Q5: Has the team evaluated its contribution well empirically, against important baselines?** e.g. no evaluation at all

+ **1 pts** **Q5: Has the team evaluated its contribution well empirically, against important baselines?** e.g. barely a sanity check for logical correctness of implementation, but no evaluation of task performance

✓ **+ 2 pts** **Q5: Has the team evaluated its contribution well empirically, against important baselines?** e.g. good start, but evaluated on oversimplified tasks, missing obvious baselines

+ **3 pts** **Q5: Has the team evaluated its contribution well empirically, against important baselines?** e.g. evaluated on challenging tasks against all obviously relevant baselines

+ **0 pts** **Q6: Has this project succeeded at making its target contribution, or systematically analyzed the technical reasons for not being able to do so?** See explanation in comments.

+ **1 pts** **Q6: Has this project succeeded at making its target contribution, or systematically analyzed the

technical reasons for not being able to do so?** See explanation in comments.

+ **2 pts** **Q6: Has this project succeeded at making its target contribution, or systematically analyzed the technical reasons for not being able to do so?** See explanation in comments.

✓ **+ 3 pts** **Q6: Has this project succeeded at making its target contribution, or systematically analyzed the technical reasons for not being able to do so?** [_Use submission specific adjustments to reduce points for this. **Note:** e.g. for full 3 pts: A well-designed novel algorithm failed to beat a baseline on a real-world dataset, but the team analyzed hand-picked samples that it performed better on and arrived at clearly justified hypotheses for what types of data their approach works well on.
Sidenote: The bar for this type of “analysis” will be higher for projects whose contribution is solely in analysis: rather than merely hand-pick some samples to formulate some initial hypotheses, they might, for example, show a sequential progression of experiments to produce increasingly refined hypotheses about the strengths and weaknesses of various existing approaches._]

+ **0 pts** **Q7: Has the report identified conclusions clearly?** No

✓ **+ 1 pts** **Q7: Has the report identified conclusions clearly?** Yes

✓ **+ 0 pts** **Q8: Which of the following contributions does it target? ** New application domain for an existing algorithm

✓ **+ 0 pts** **Q8: Which of the following contributions does it target? ** New or expanded publicly available dataset

+ **0 pts** **Q8: Which of the following contributions does it target? ** New publicly available code implementation of any algorithm (different from previous available implementations, if any)

+ **0 pts** **Q8: Which of the following contributions does it target? ** New ML algorithm or technique

✓ **+ 0 pts** **Q8: Which of the following contributions does it target? ** Analysis leading to improved understanding of existing algorithms

+ **0 pts** **Q8: Which of the following contributions does it target? ** Other, pre-approved

Very Good project. Well done team!
The results section is a bit weak - consolidated results have not been supplied. While GRU based SOLO did not perform well, a comparison table against SOLO (RESNET-50 backbone) and SOLO (unmodified) could have been supplied for the metrics under consideration.
I noticed that some values are mentioned in the abstract but even there, the hyperparameters used to generate them are not clear.

SOLO in Biomedic Image and Exploration

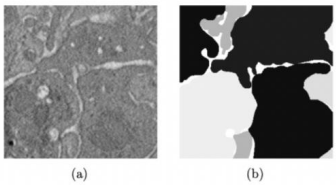
Team: Yunqi Xu, Yuxin Wang, Zhenglin Zhang. Project Mentor TA: Swati

1) Abstract

Pathological formation of hemostatic plugs is often a crucial part of fatal diseases, while the current studies rely heavily on manually annotated platelets on the scanning of the slices of the plug. In this project, we explore the combination of traditional biomedical imaging dataset with state-of-the-art instance segmentation models. From moderated compressed dataset created with commercial software, we first use the DBSCAN algorithm to convert them to standard data format that is compatible with the usage of data in the computer vision field. We then modified the SOLO with a customized ResNet50 backbone for a binary classification problem, and searched a few hyperparameters to tune the model from COCO dataset to biomedical dataset. The average precision of the four sets of hyperparameters are 0.473, 0.584, 0.647 and 0.461. We also make a few attempts to add GRU for FPN output levels, but due to the extra complexity that it brings, none of them succeeds. A brief video presentation of the project can be found in the link included in the footnote.

2) Introduction

Hemostatic plug, also known as the platelet plug, is an aggregation of platelets forming around injuries of the blood vessel wall as a part of the hemostasis mechanism. The platelets in the blood scream are activated by the exposure to the non-blood vessel environment, whose adhesive building up forms a preliminary protection against external contaminants as well as further blood loss[1]. However, pathological formation of hemostatic plugs is often a crucial part of fatal diseases. For example, in hemophilia and von Willebrand disease, the patients' platelets fail to produce fibrin that is essential to the stabilization of the plug. As a consequence, the plug breaks down and exposes the injuries[2]. Such mechanism brings the need for early diagnosis of pathological formations of the plug from patterns of their scanning. While the current studies of hemostatic plug formation rely heavily on manually annotated platelets on the scanning of the slices of the plug (as shown in figure 1), in this project, we explore an automatic annotation and separation of cells from the sliced scanning of the plug using the SOLO[3] without the burden of traditional segmentation algorithms that cannot be implemented in parallel fashion. The one-shot prediction of individual cell masks also allows the incorporation of spatial prior information, where the channels of images are continuous scanning of slices of tissue, or the evolution of the cells in time scale, that leads to the 3-D point clouds reconstruction of individual cells with an additional z−axis.The inputs are original microscopic images of tissues (x,y) and the outputs are (y). We divide the image into a grid of 7×7 cells (suboriginal), resulting in S2 classes of center locations. We mask and label these images manually to create submasks and sub-labeled images. The suborigine is x, and the submasks are y in SOLO model input. From the ground truth masks, we use density based clustering methods to recover masks for individual cells. The images and individual masks are then fed into the SOLO as training data. We use the average precision to estimate the accuracy of the prediction, as well as relative L2 error of the merged masks as opposed to the ground truth masks. Other tasks can only be performed on static images. We perform SOLO on dynamic images which significantly improves the accuracy and efficiency of biological image analysis.



(a) (b)

Figure 1: A visualization of existing data. (Left) Original scanning. (Right) Manually annotated mask.

3) Background

Echoing the vibrant community of deep learning, the computer vision researchers are presenting an unprecedentedly rapid development of instance segmentation methods that efficiently detect the presence of objects from multiple categories in images without human intervention. While achieving considerable success in Common Object in Context (COCO) dataset[4], the state of the art instance segmentation models, such as SOLO [3], SSD [5], and Mask R-CNN [6], are rarely seen in the fields of biomedical imaging, where U-Net based methods dominate especially for cell detection and identification in the microscopic tissue images. On the contrary to the popular instance segmentation models mentioned above, where the masks are predicted after the determination of their locations,

these U-Net based methods predict semantic segmentation mask first for the entire image, before using algorithms such as watershed or regional proposal network [7] [8] [9] to identify masks for individual objects in the image. The inverted two-stage detection fails to anchor the masks at their locations and therefore compromising the ability to trace individual objects across the images. Meanwhile, region-based Convolutional Network method (R-CNN) [10] and its descendants, e.g. Fast R-CNN [11] and Faster R-CNN [12], contribute a popular diagram for object detection. They use regional proposals or bounding boxes to localize and segment the target. Following the work of Faster R-CNN, Mask R-CNN [6] extends this method to instance segmentation tasks by adding an extra branch of mask prediction besides the bounding box prediction branch as an end-to-end solution. It's easy to be trained and regularized. Later, Wang, X.et al.[3] introduced Segmenting Objects by Locations (SOLO). By assigning categories on pixel level via the instance's spatial information (location and size), it successfully transforms the traditional instance segmentation to a one-shot classification-solvable task, achieving on par performance with Mask R-CNN but with a simpler one-shot architecture.

4) Summary of Our Contributions

1. Contribution(s) in Code: In original SOLO paper, they use the ResNet-101 and FPN to extract feature maps and train the model. But compared with their COCO dataset, our dataset is smaller and we do not need such a 'deep' network. So we replace ResNet-101 with ResNet-50 and combine it with 4 layers of FPN to make our new SOLO model. Besides using this ResNet architecture we also feed data to GRN and compare their results.
2. Contribution(s) in Application: Traditional machine learning or deep learning algorithms could only detect and tell a still image, SOLO developed based on YOLO allows for real-time pixel-level object recognition. Instead of using it in our daily life, using SOLO for cell identification in medicine greatly reduces human error in cell analysis.
3. Contribution(s) in Data: In SOLO paper, the authors used COCO dataset for common object detection. But we collect our own cell data from our Penn Medical School and these data are real individual tissues, which makes our model more realistic to use.

5) Detailed Description of Contributions

Contribution(s) in Code: The code contains 4 files.
config.py: It contains some basic parameters we will use in the training part.
data_loader.py: it used to pre-process original data, output masked and labeled data and split the data to train and test dataset.
architecture.py: It contains architectures we used in this project (GRU, FPN, Botteleneck, SOLO head)
solo_train.py: we use it to train the model and show the related loss.
1.SOLO Achitecture
The SOLO model's architecture shown in Fig.2. The central idea of the SOLO framework is to reformulate the instance segmentation as two simultaneous categories - aware prediction problems, which are Semantic category and Instance Mask. Given an input image, we first use FCN to get the feature map of it. In the original paper, they use the ResNet-101, while we use the ResNet-50 FPN instead. And then input these feature maps into these two simultaneous prediction problems (called SOLO head architecture).

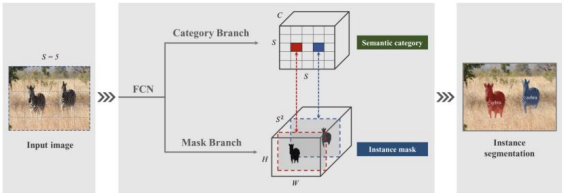


Fig.2. SOLO Framework

2.Resent-50 FPN
FPN backbone：We use Resnet50 + 4 layers Feature Pyramid Network as the backbone of SOLO. The difference we made about the regular Resnet is that we chose 4 FPN layers so that the first feature map produced by the first residual block should also be taken into consideration and parameters need to be adjusted. The detail of the process is similar to regular FPN: each layer comes after the residual block, after upsampling and soothing then blends with the lower layer, the final feature maps of each layer are used to predict different sizes of masks.
Resent-50 FPN Architecture: The ResNet-50 FPN architecture is shown in Fig.3. The top part of figure is ResNet-50 which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer

and the input image of size H*W is processed to extract feature maps of different sizes. The bottom part of figure 2 depicts the FPN architecture, it provides a top-down pathway to construct higher resolution layers from a semantic rich layer and further processes the last three states from the ResNet-50 Network. In this way, we could get a more effective feature map to detect the different object sizes.

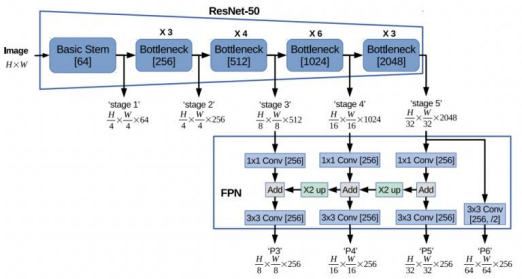


Fig.3. ResNet-50 FPN architecture[1]

SOLO Head: The feature map we get from ResNet-50 FPN would be the input of our SOLO head architecture. (Shown in Fig.4.) In the semantic category prediction branch, the SOLO will predict C-dimensional output to indicate the semantic class probabilities and the output space is S×S×C. The instance mask branch is parallel with the semantic category prediction and each positive grid cell will generate the corresponding instance mask.

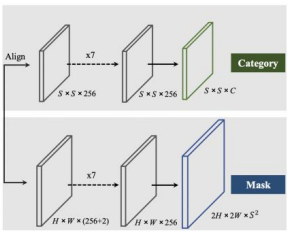


Fig.4. SOLO head architecture

The one to one correspondence is established between the semantic category and class-agnostic mask by the following equation:

$$k = i \cdot S + j \tag{1}$$

As a binary classification task, the mask loss function is simplified from focal loss to binary cross entropy.

3.GRU

Recurrent structure：To utilize the spatial information provided in the sliced scanning, we use a recurrent structure that combines features from nearby slices. Among sequence processing models, Gated Recurrent Units (GRU) [14] has been a popular light-weighted alternative to the older Long Short-Term Memory (LSTM)[15]. Empirically, it has better performance on certain smaller and rare datasets. In 2015, Bi-LSTM [16] was proposed by Huang et al. as a bidirectional variant of LSTM that utilizes both the past and future information. We will similarly construct a bidirectional variant of GRU, where the linear layers are replaced by a convolutional layer with kernel size 3 × 3. Instead of doubling the hidden dimension, we first concatenate the hidden state from both directions together, and then apply a 1 × 1 convolution to recover the original dimension. Figure 3 shows a schematic drawing of the architecture with recurrent units.

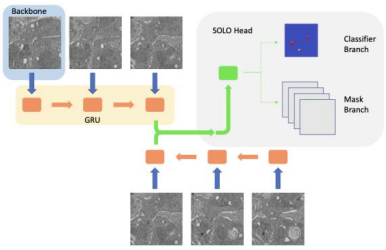


Fig.5. Architecture of SOLO with GRU

GRU：Gate Recurrent Unit is one kind of Recurrent Neural Network. Compared with LSTM, GRU only has a reset gate and update gate that decide what information should be passed to the output. We calculate the updated gate for time step t using the formula:

$$Z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1}) \tag{2}$$

The reset gate is used to decide how much of the past information to forget. The formula is shown below.

$$r_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1}) \tag{3}$$

After using the reset gate to store the relevant information from the past, we get the current information shown below. Where W, U is the related weight, h_{t-1}holds the information for the previous t-1 units

$$h'_t = \tanh\left(Wx_t + r_t \odot Uh_{t-1}\right) \tag{4}$$

We then use the updata gate to combine the previous content and current content, In this formula, we apply element-wise multiplication to them.

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h'_t \tag{5}$$

Since we have to detect the cell in a real-time video, GRU perfectly solves this 'time' problem, helping us to store and filter the information using their update and reset gates.

5.1 Methods

Dataset description：As shown in 1, the dataset comprises 69 original scanning of the hemostatic plug with size 1250 × 1250, and manually annotated mask where the area of the platelets are labeled with different gray scales or RGB colors. We first crop each image to 16 images with size 313 × 313 to reduce the size and increase the number of samples. Among these 1104 images, we use 80 − 20 splits for the training set and testing set. We are requested by the data providers to not disclose the dataset.

Dataset preprocessing：To overcome such anomaly from compression and extract individual masks from the manually labeled mask for entire images, we use Density-based spatial clustering of applications with noise (DBSCAN)[13] to identify clusters of pixels (i.e., individual platelet masks) with close grayscale or RGB color values. DBSCAN classifies points as core points, boundary points, and noise points from the number of in their ε-neighborhood. Then core points that are close to each other, as well as boundary points that are close to the core points of the same labels, are identified as a cluster. Specifically, we use the Euclidean distance measured together on the positional indices and color space as a distance metric for the pixels. We use ε = 2 to define the neighborhood of a point, and N = 9 to define how many points in the neighborhood that a point must have in order to be a core point. Figure 2 shows a sample extraction of individual masks.

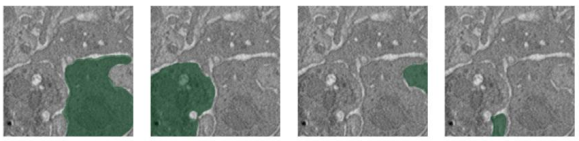


Fig.6. Examples of individually extracted mask

Once the masks are obtained, the bounding box parameters are chosen as 0.1% to 99.9% quantiles at each dimension to avoid lingering pixels.

For target assignment to FPN feature levels, we follow the routine in the original SOLO paper, and test a variety of combinations of scale ranges and feature grid sizes. However, since we have limited knowledge of the dataset, our choices of these hyperparameters may not be optimal.

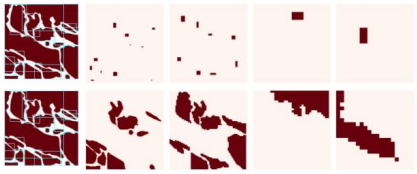


Fig.7. A visualization of target assignments. On the left is a picture of masks bounded by their bounding boxes.(Top) Activated pixels on the feature map in each layer. (Bottom) The corresponding masks assigned to the activated pixels.

Post-processing: For each image, we choose top 200 scored masks whose categorical score exceeds 0.5. Then their confidence score, a multiplication of the categorical and maskness, are used for the matrix Non-Maximal Suppression (NMS), after which top 15 masks are output as final prediction.

Evaluation metrics: Mean Average Precision (mAP) is a popular metric in the object detection domain. By computing the average precision of each class detection with the help of IoU threshold and then averaging it by the number of classes, we can know how well the model performs.

Training We choose: Adam optimizer [17] for its fast convergence. Although heuristic circulates that the SGD with momentum may promote better generalization, our pursuit to higher test performance was hindered by the limited time. We choose the learning rate as 1e − 3 as standard practice, and have annealing schedule per 15 epoch by a factor of 0.95. The loss function is designed as

$$L = \lambda_c L_{cate} + \lambda_m L_{mask} \tag{6}$$

Lcate is the binary cross entropy loss for the classifier head that distinguish a feature grid from background, and Lmask is the mask loss that can be expressed as

$$L_{mask} = \frac{1}{N_{pos}} \sum d_{mask}\left(p^i, q^i\right) \tag{7}$$

where pi, qi are the assigned mask, and the predicted mask, respectively. And dmask, the dice loss, is defined as

$$d_{mask} = 1 - \frac{2\langle p, q\rangle}{\langle p, p\rangle + \langle q, q\rangle} \tag{8}$$

where $\langle \cdot, \cdot \rangle$ is the inner product on the flattened masks.

We use λc = λm = 1 as the weight. Figure 8 shows the descending of loss curves during the training for a successful trial. Although the testing classification loss increases, we consider that the overfitting is also a necessary part for the reconstruction of the training set, and we therefore continue training. The test mask loss stays almost constant during the training.

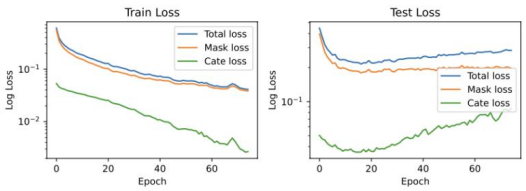


Fig.8. Loss curves of a successful training of SOLO

5.2 Experiments and Results

Inference result: Figure 9 shows sample outputs from train and test images. Due to the page limit, we cannot provide an exhausting list of post-processed predictions.

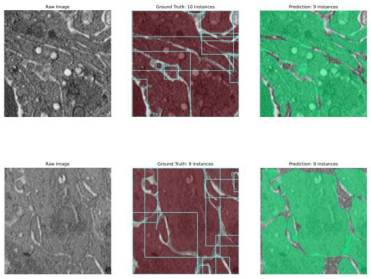


Fig.9. Sample post processed inference results from SOLO without GRU.(TOP) On one of the train images.(Bottom) on one of the test images

For SOLO with GRU, unfortunately, the model fails to learn and does not detect any object due to the extremely low confidence score returned by the model. Both the implementation given by earlier studies and our own fail for the same reason. Figure 10 shows the loss curves during the training that fail to decrease. We will therefore not provide any result for SOLO with GRU, but we will include a section for the potential cause and fix for the failure in the discussion section later.

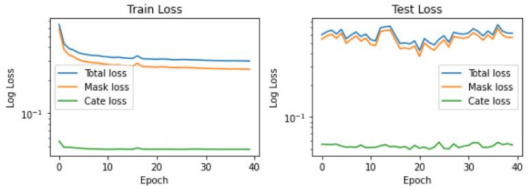


Fig.10. Loss curves of SOLO with GRU fail to decrease.

Average precision For each architecture, we use 4 fixed sets of scale ranges for target generation to test

6) Compute/Other Resources Used

All the computations are run on a single NVIDIA RTX 2080 graphic card with 8Gb memory. Table shows some general computational cost of the chosen model.

Table.1. model overview

Model	Evaluation	Post-Inference	Time per epoch
SOLO	44 ± 8.52 ms	197 ± 27.8 ms	~ 160 s
SOLO with GRU	133 ± 15 ms	508 ± 142 ms	~ 320 s

Computation cost of evaluation (a batch of 4), inference with post processing (single image), and a training epoch (~ 820 images).

7) Conclusions

SOLO with GRU fails to learn. However, this result is still within reasonable expectation since the field convolutional recurrent neural network in general has only limited studies. While the convolution itself is an analog of linear operation on a regular grid instead of a vector, whether we can replace the linear operation in recurrent structure with convolution is still unknown. And in literature, researchers also admitted that the introduction of recurrent structure may not always bring at least comparable results.

Besides the architecture, the errors in manual annotation also accumulate and bring conflicts to scanning of even nearby slices, due to the fact that humans are not too capable of recognizing spatial correlation, especially when the images to be labeled are not fed with spatial order. Such discrepancy in the images data makes the introduction of GRU as well as additional complexity far less economic. However, we do believe that based on the intuition that larger cells are more likely to have significant spatial correlation, we can apply the recurrent architectures to the last few FPN layers which are assigned to detect those larger objects.

(Exempted from page limit) Other Prior Work / References (apart from Sec 3) that are cited in the text:

[1] J. Ware and Z.M. Ruggeri. Platelet adhesion receptors and their participation in hemostasis and thrombosis. *Drugs of Today*, 37(4):265, 2001. ISSN 1699-3993. doi:10.1358/dot.2001.37.4.620592. URL http://journals.prous.com/journals/servlet/xmlxsl/pk_journals.xml_summary_pr?p_JournalId=4&p_RefId=620592&p_IsPs=N.

[2] Northwestern Medicine. Blood Clot Disorder. URL https://www.nm.org/conditions-and-care-areas/hematology/blood-clot-disorder.

[3] Xinlong Wang, Tao Kong, Chunhua Shen, Yuning Jiang, and Lei Li. Solo: Segmenting objects by locations. In *European Conference on Computer Vision*, pages 649–665. Springer, 2020.

[4] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015.

[5] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[8] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

[9] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38 (1):142–158, 2015.

[11] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[12] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 201, 2015.

[13] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.

[14] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[15] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997. ISSN 0899-7667. doi:10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8. 1735.

[16] Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*, 2015.

[17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

Broader Dissemination Information:

Your report title and the list of team members will be published on the class website. Would you also like your pdf report to be published?
NO

If your answer to the above question is yes, are there any other links to github / youtube / blog post / project website that you would like to publish alongside the report? If so, list them here.

- <description in a few words>: <URL>
- <description in a few words>: <URL>
  ...

(Exempted from page limit) **Work Report: This may look like your GANTT chart from the midway report, with more completed steps now. Okay to modify.** (Mark completed steps in green, as shown here. For convenience, you may split into two charts, one till Nov 8, and another for after Nov 8, placed one below the other.)

PERSON (S)	TASK (S)	Wk5				Wk6				Wk7				Wk8				Wk9		
		OCT																NOV		
		S 3	M 4	W 6	Th 7	S 10	M 11	W 13	Th 14	S 17	M 18	W 20	Th 21	S 24	M 25	W 27	Th 28	S 31	M 1	W 3	Th 4
Yuxin Wang	Cutting each figure into 5*5 subfigures in order																			
Yunqi Xu	Label the figure																			
Yuxin Wang, Yunqi Xu	Solo head																			
Zhenglin Zhang	Resnet+FPN backbone																			
Zhenglin Zhang	MAP performance																			
Yuxin Wang, Yunqi Xu, Zhenglin Zhang	Final Report																			
...	Task 7																			
...	Task 8																			
...	Task 9																			
...	Task 10																			

(Exempted from page limit) Attach your midway report here, as a series of screenshots from Gradescope, starting with a screenshot of your main evaluation tab, and then screenshots of each page, including pdf comments. This is similar to how you were required to attach screenshots of the proposal in your midway report.

7 / 7 pts

QUESTION 1

Evaluation Question [Select all pages] **7** / 7 pts

✔ + 1 pt	Does the report follow the provided template including the 4-page limit (excluding exempted portions), with reasonable responses to all questions?
✔ + 2 pts	Has feedback from the last round been effectively addressed?
✔ + 1 pt	Has the team identified a clear topic and viable new target contribution, as per the project specifications provided in class?
✔ + 1 pt	Has the team moved in a non-trivial way towards their target contribution?
✔ + 2 pts	Has a clear and systematic work plan been formulated for the remaining weeks?

1) The feedback is mostly addressed, but I still find that the code contribution part is a bit lacking. You plan to replace RESNET-101 with RESNET-50 and keep the rest of the architecture from the original paper/s, which is a logical choice given the smaller size of the dataset. But there is no "novelty" here, so I would advise that you try and implement the GRU part as well (which you have not promised currently).
2) Since GPU and train time seems to be a concern, I advise you to not train your RESNET from scratch, initialize the layers with pre-trained weights publicly available. This will help speed up things and improve performance.
3) Have you kept a separate test set portion (train-test split?). If not, make sure to do so and report your results on the separated test set in analysis section. Also, since hyperparameter tuning seems to be a concern, you can also keep a small validation set and run some experiments on it to narrow down the hyperparameter space.
4) Risk mitigation plan is a bit weak. If you end up not changing the original architecture at all as you have suggested based on time constraints, the code contribution part will be completely gone, so you will end up losing points. Try to quicken up the pace and implement as much as you can!

Overall, good progress! (seems like you are done with the data preprocessing part of it).
Do address the feedback above and focus on the ML model side of things now. Feel free to set up a meeting if needed.
Keep it up, and all the best for the final report!

SOLO in Biomedic Image and Exploration

Team: Yunqi Xu, Yuxin Wang, Zhenglin Zhang. **Project Mentor TA: Swati**

1) Introduction

The inputs are original microscopic images of tissues (x,y) and the outputs are (y). We divide the image into a grid of 7×7 cells (suboriginal), resulting in S^2 classes of center locations. We mask and label these images manually to create submasks and sub-labeled images. The suboriginal would be x, and the submasks would be y in SOLO model input. From the ground truth masks, we will use density based clustering methods to recover masks for individual cells. The images and individual masks are then fed into the SOLO as training data.

We will use the average precision to estimate the accuracy of the prediction, as well as relative L2 error of the merged masks as opposed to the ground truth masks.

Other tasks can only be performed on static images. We aim to perform SOLO on dynamic images which will significantly improve the accuracy and efficiency of biological image analysis.

2) How We Have Addressed Feedback From the Proposal Evaluations(Yulia)

The detailed of dataset:

The data we collected is from Penn Medical School, and the original data is a 8s's video of cell.(https://drive.google.com/file/d/1Sfsfize7OD7RBHGMhvByPofJD4uzEG6z/view?usp=sharing). We cut the video into 100 cell images shown below, and each image is 1250*1250. After that, we split the original image into small 7*7 image (called suboriginals), the pixel size of each 'suboriginals' is 313*313. We mask and label these images by hand to create submasks and sublabeled image. The suboriginals would be x, and the submasks would be y in our SOLO model input.

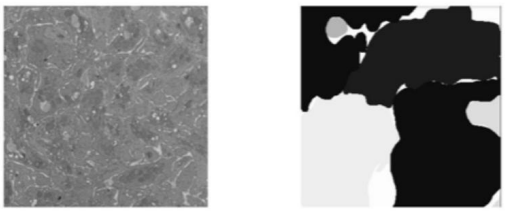


Fig.1. Original image and labeled image

The details of Architecture:

The main architecture of our model consists of two main parts: ResNet-50 FPN and prediction head.(The detailed content could be found in 5) contributions in code). We use ResNet-50 FPN (Feature Pyramid Network) to generate a pyramid of feature maps and use it as input for each prediction head: semantic category and instance mask.

3) Prior Work We are Closely Building From

A.Title: SOLO: Segmenting Objects by Locations (URL: https://arxiv.org/abs/1912.04488. Code link:https://github.com/WXinlong/SOLO)
This paper transforms instance segmentation into a classification problem by introducing the concept of "instance class", which assigns a class to each pixel in an instance based on its location and size. Compared to Mask R-CNN, the architecture is simpler but more effective and is a one-stage instance segmentation method.
B.Title: Feature Pyramid Networks for Object Detection (URL: https://arxiv.org/abs/1612.03144. Code link: https://github.com/DetectionTeamUCAS/FPN_Tensorflow.)
Most of the original object detection algorithms only use top-level features for prediction, but the low-level feature semantic information is small, but the target position is accurate. The high-level feature semantic information is rich, but the target position is rough. Although some algorithms use multi-scale feature fusion, they generally use the fused features to make predictions. The difference in this paper is that the predictions are performed independently at different feature layers.

4) What We are Contributing

1. Contribution(s) in Code: Replacing ResNet-101 with ResNet-50 to fit a small dataset of tissues. (we may also consider using GRU if we have enough time during the final period)
2. Contribution(s) in Application: Using SOLO in the medical field to achieve real-time pixel-level cell detection.
3. Contribution(s) in Data: Collection data from a real individual tissual video inPenn Medical School.
4. Contribution(s) in Algorithm: N/A
5. Contribution(s) in Analysis: N/A

5) Detailed Description of Each Proposed Contribution, Progress Towards It, and Any Difficulties Encountered So Far

- Contributions in Code: In original SOLO paper, they use the ResNet-101 and FPN to extract feature maps and train the model. But compared with their COCO dataset, our dataset is smaller and we do not need such ‘deep’ network. So we replace ResNet-101 with ResNet-50 and combine it with FPN to make our new SOLO model.

The SOLO model’s architecture shown in Fig.2. The central idea of the SOLO framework is to reformulate the instance segmentation as two simultaneous categories - aware prediction problems, which are Semantic category and Instance Mask. Given an input image, we first use FCN to get the feature map of it. In the original paper, they use the ResNet-101, while we use the ResNet-50 FPN instead. And then input these feature maps into these two simultaneous prediction problems (called SOLO head architecture).

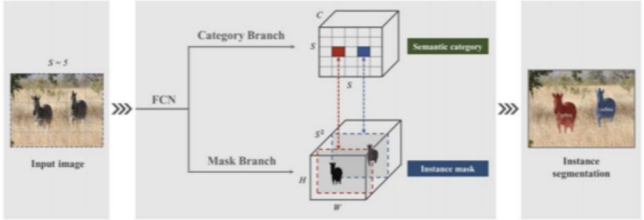


Fig.2. SOLO Framework

The ResNet-50 FPN architecture is shown in Fig.3. The top part of figure is ResNet-50 which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer and the input image of size H*W is processed to extract feature maps of different sizes. The bottom part of figure 2 depicts the FPN architecture, it provides a top-down pathway to construct higher resolution layers from a semantic rich layer and further processes the last three states from the ResNet-50 Network. In this way, we could get a more effective feature map to detect the different object sizes.

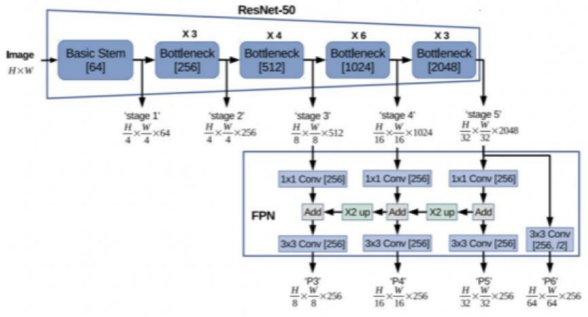


Fig.3. ResNet-50 FPN architecture[1]

The feature map we get from ResNet-50 FPN would be the input of our SOLO head architecture. (Shown in Fig.4.) In the semantic category prediction branch, the SOLO will predict C-dimensional output to indicate the semantic class probabilities and the output space is S×S×C. The instance mask branch is parallel with the semantic category prediction and each positive grid cell will generate the corresponding instance mask.

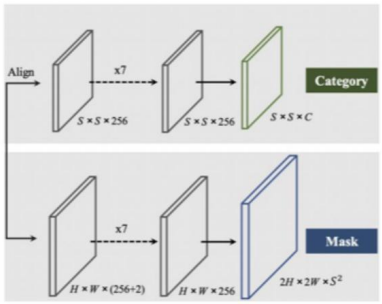


Fig.4. SOLO head architecture

The one to one correspondence is established between the semantic category and class-agnostic mask by the following equation:

$$k = i \cdot S + j \tag{1}$$

The SOLO is then followed by the training loss function to optimize the model. The training loss function is shown below:

$$L = L_{\text{cate}} + \lambda L_{mask} \tag{2}$$

L_{cate} is the conventional Focal Loss for semantic category classification and L_{mask} is the loss for mask prediction:

$$L_{mask} = \frac{1}{N_{pos}} \sum_{k} \mathbb{1}_{\{\mathbf{p}^*_{i,j}>0\}} d_{mask}\left(\mathbf{m}_k, \mathbf{m}^*_k\right) \tag{3}$$

- Contributions in Application: Traditional machine learning or deep learning algorithms could only detect and tell a still image, SOLO developed based on YOLO allows for real-time pixel-level object recognition. Instead of using it in our daily life, using SOLO for cell identification in medicine greatly reduces human error in cell analysis.
- Contributions in Data: In SOLO paper, the authors used COCO dataset for common object detection. But we collect our own cell data from our Penn Medical School and these data are real individual tissues, which makes our model more realistic to use.

5.1 Methods

We use Density-based spatial clustering of applications with noise (DBSCAN) to identify clusters of pixels. DBSCAN classifies points as core points, boundary points, and noise points from the number of in their ε-neighborhood. Then core points that are close to each other, as well as boundary points that are close to the core points of the same labels, are identified as a cluster. Specifically, we use the Euclidean distance measured together on the positional indices and color space as a distance metric for the pixels. We use ε = 2 to define the neighborhood of a point, and N = 9 to define how many points in the neighborhood that a point must have in order to be a core point.

For target assignment to FPN feature levels, we follow the routine in the original SOLO paper, and test a variety of combinations of scale ranges and feature grid sizes. However, since we have limited knowledge of the dataset, our choices of these hyperparameters may not be optimal.

5.2 Experiments and Results

Our key question of the experiment is trying to identify whether each pixel belongs to the real tissue or not, so this is a semantic pixel segmentation problem. The baseline follows the original SOLO paper and we combine it with GRU to find if the result will get better(not promised). Obviously the performance metrics are going to be the mean average precision/average precision, of which the true positives are those IOU of two masks bigger than the threshold.

We haven't completed any of these performance evaluations yet because these need to be done after all the results come out at last.

6) Risk Mitigation Plan

We won't start with a simplified setting unless we don't have enough time for the regular SOLO training with a simplified backbone. We will directly use the original SOLO paper with Resnet 101 and this will not decrease the accuracy of the results for more layers due to the feature of Resnet only more running time.

Also, less time means the absence of SOLO with GRU. However, our model will never "fail" for the regular SOLO model but low mean average precision at most. SOLO with GRU might fail to return the high-confidence results but this is only an extension.

(Exempted from page limit) Other Prior Work / References (apart from Sec 3) that are cited in the text:

1. Author 1, Author 2, Author 3, “Deep Learning for predicting cat moods as a function of TV watching”, SIGBOVIK 2020.Panero Martinez, R.; Schiopu, I.; Cornelis, B.; Munteanu, A. Real-Time Instance Segmentation of Traffic Videos for Embedded Devices. *Sensors* **2021**, *21*, 275.
2. Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.

(Exempted from page limit) Supplementary Materials if any (but not guaranteed to be considered during evaluation):

1 Grading [select all pages for this question] **14 / 15**

+ **0 pts** **Q1. Does the final report match the provided template?** No

✓ **+ 1 pts** **Q1. Does the final report match the provided template?** Yes

+ **0 pts** **Q2. To what extent has feedback from previous rounds been effectively addressed?** e.g. Ignored Completely

+ **1 pts** **Q2. To what extent has feedback from previous rounds been effectively addressed?** e.g. barely addressed anything substantial

+ **2 pts** **Q2. To what extent has feedback from previous rounds been effectively addressed?**e.g. clear response to feedback, but ignored some key feedback

✓ **+ 3 pts** **Q2. To what extent has feedback from previous rounds been effectively addressed?** e.g. addressed all key feedback from both previous rounds

+ **0 pts** **Q3. To what extent does this report identify and target a novel contribution, which it sufficiently differentiates from prior work** e.g. clearly no novel contribution at all,

+ **1 pts** **Q3. To what extent does this report identify and target a novel contribution, which it sufficiently differentiates from prior work**e.g. some novelty likely, but no clear differentiation within the report.

✓ **+ 2 pts** **Q3. To what extent does this report identify and target a novel contribution, which it sufficiently differentiates from prior work** e.g. clearly novel contribution, well differentiated from prior work

+ **0 pts** **Q4: Did the team execute a contribution-focused, step-by-step work plan, sharing workload among team members, mitigating obvious risks, as needed?** e.g. no clear plan at all, or barely any execution.

+ **1 pts** **Q4: Did the team execute a contribution-focused, step-by-step work plan, sharing workload among team members, mitigating obvious risks, as needed?** e.g. good plan, but limited progress towards execution

✓ **+ 2 pts** **Q4: Did the team execute a contribution-focused, step-by-step work plan, sharing workload among team members, mitigating obvious risks, as needed?** e.g. good plan, and executed well and almost to completion

+ **0 pts** **Q5: Has the team evaluated its contribution well empirically, against important baselines?** e.g. no evaluation at all

+ **1 pts** **Q5: Has the team evaluated its contribution well empirically, against important baselines?** e.g. barely a sanity check for logical correctness of implementation, but no evaluation of task performance

✓ **+ 2 pts** **Q5: Has the team evaluated its contribution well empirically, against important baselines?** e.g. good start, but evaluated on oversimplified tasks, missing obvious baselines

+ **3 pts** **Q5: Has the team evaluated its contribution well empirically, against important baselines?** e.g. evaluated on challenging tasks against all obviously relevant baselines

+ **0 pts** **Q6: Has this project succeeded at making its target contribution, or systematically analyzed the technical reasons for not being able to do so?** See explanation in comments.

+ **1 pts** **Q6: Has this project succeeded at making its target contribution, or systematically analyzed the technical reasons for not being able to do so?** See explanation in comments.

+ **2 pts** **Q6: Has this project succeeded at making its target contribution, or systematically analyzed the technical reasons for not being able to do so?** See explanation in comments.

✓ **+ 3 pts** **Q6: Has this project succeeded at making its target contribution, or systematically analyzed the technical reasons for not being able to do so?**

[_Use submission specific adjustments to reduce points for this. **Note:** e.g. for full 3 pts: A well-designed novel algorithm failed to beat a baseline on a real-world dataset, but the team analyzed hand-picked samples that it performed better on and arrived at clearly justified hypotheses for what types of

data their approach works well on.
Sidenote: The bar for this type of "analysis" will be higher for projects whose contribution is solely in analysis: rather than merely hand-pick some samples to formulate some initial hypotheses, they might, for example, show a sequential progression of experiments to produce increasingly refined hypotheses about the strengths and weaknesses of various existing approaches._]

+ **0 pts** **Q7: Has the report identified conclusions clearly?** No
✓ **+ 1 pts** **Q7: Has the report identified conclusions clearly?** Yes
✓ **+ 0 pts** **Q8: Which of the following contributions does it target? ** New application domain for an existing algorithm**
✓ **+ 0 pts** **Q8: Which of the following contributions does it target? ** New or expanded publicly available dataset**
+ **0 pts** **Q8: Which of the following contributions does it target? ** New publicly available code implementation of any algorithm (different from previous available implementations, if any)
+ **0 pts** **Q8: Which of the following contributions does it target? ** New ML algorithm or technique
✓ **+ 0 pts** **Q8: Which of the following contributions does it target? ** Analysis leading to improved understanding of existing algorithms**
+ **0 pts** **Q8: Which of the following contributions does it target? ** Other, pre-approved

Very Good project. Well done team!
The results section is a bit weak - consolidated results have not been supplied. While GRU based SOLO did not perform well, a comparison table against SOLO (RESNET-50 backbone) and SOLO (unmodified) could have been supplied for the metrics under consideration.
I noticed that some values are mentioned in the abstract but even there, the hyperparameters used to generate them are not clear.