CV23S Ganzin: Pupil Tracking Final Project Presentation

第17組　越共貪啃奇

Rank 6 out of 19

謝承恩	蔣沅均	江讀晉	蕭承瀚
R11943015	R11921041	R11921038	R11424006

Outline

- Problem formulation
- Framework
- Related works
- Method
- Experiment
- Conclusion

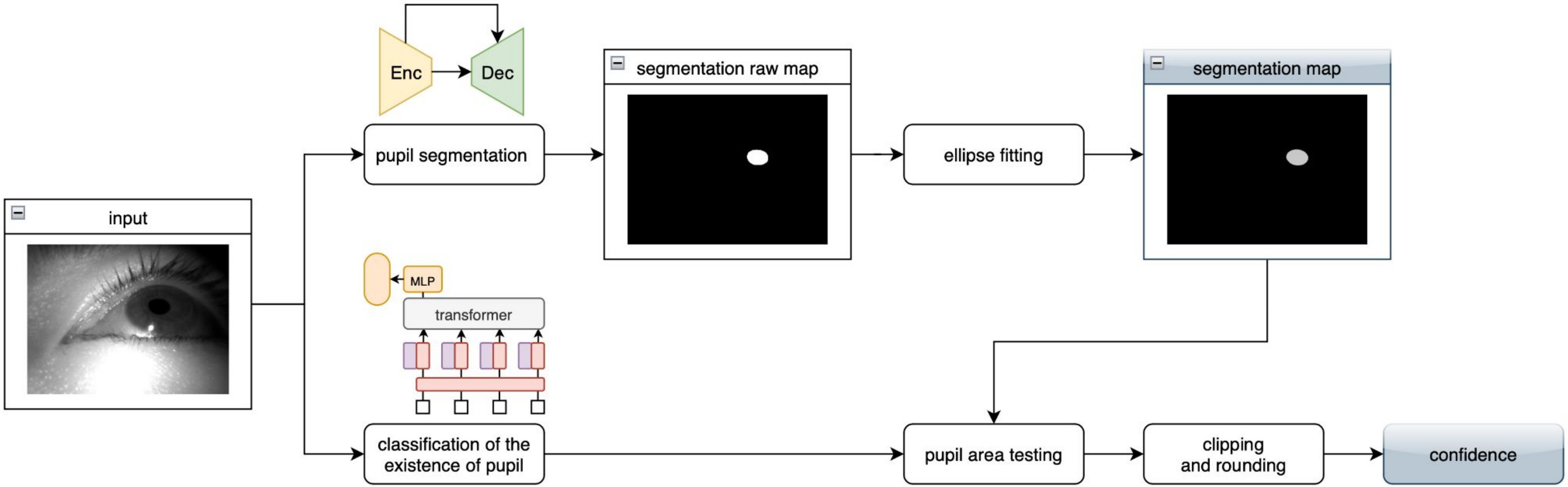
Problem formulation

- Two goals
  - Segment pupils in eye images
  - Determine the existence of pupils
- Our proposed solution
  - Specialized semantic segmentation task
    - The segmentation task requires the segmentation of the entire pupil, even when it is partially occluded. To address this issue, specific loss functions are required to handle such cases.
  - A two-class classification task aimed at determining the presence pupils.
    - To address the issue of an imbalanced dataset, we introduce Masked Autoencoder (MAE) as a self-supervised training strategy. MAE enables the model to learn robust visual representations during pre-training.

Problem solving

- Initially, we relied solely on the rules from the segmentation map to generate confidence files.
    - Unfortunately, this approach yielded unsatisfactory results.
    - Therefore, we decided to train an additional classifier.
- To the best of our knowledge, we chose a masked autoencoder-pretrained ViT-L.
    - ViT-L successfully learn how to extract features from images.
- In summary, our approach used two separate models for segmentation and confidence prediction respectively. Then, we integrated the information from both models using post-processing techniques.

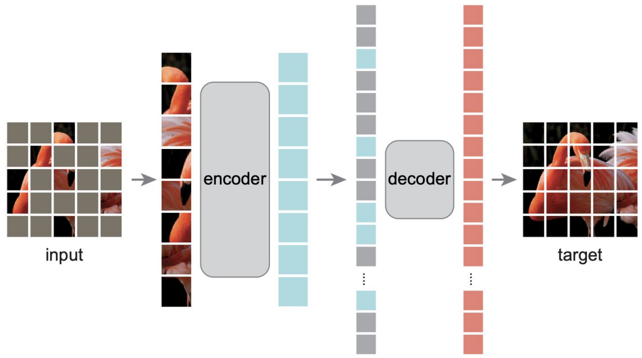
Framework


Enc
Dec
pupil segmentation
segmentation raw map
ellipse fitting
segmentation map
input
MLP
transformer
classification of the existence of pupil
pupil area testing
clipping and rounding
confidence

Related work — segmentation

- EllSeg: An Ellipse Segmentation Framework for Robust Gaze Tracking
    - Author Aayush K.Chaudhary, Rakshit Kothari, Manoj Acharya, *et.al.*
    - Jounal 2021 IEEE Transactions on VCG
    - Affiliation Rochester Institute of Technology, USA
- EllSeg enables prediction of the pupil as *full elliptical structures* despite the presence of occlusions.

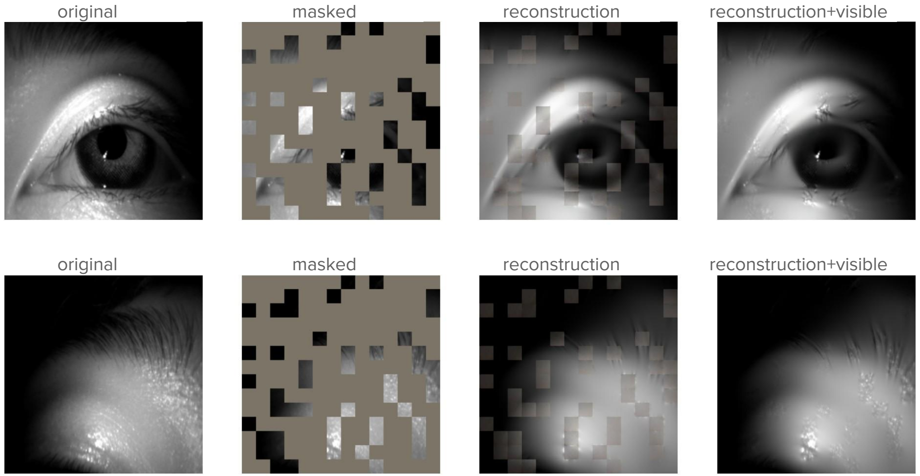
Related work — classification



- Masked autoencoder (MAE)
  - Author Kaiming He, Xinlei Chen, *et.al.*
  - Conference 2022 CVPR
  - Affiliation Facebook AI Research
- MAE allows for learning high-capacity models that generalize well.
- Self-supervised learning has demonstrated its superiority in learning robust features, thereby enhancing knowledge transfer to downstream tasks.

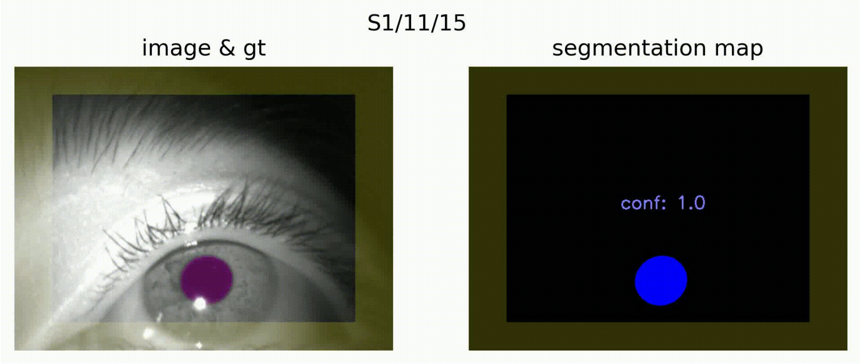
Method

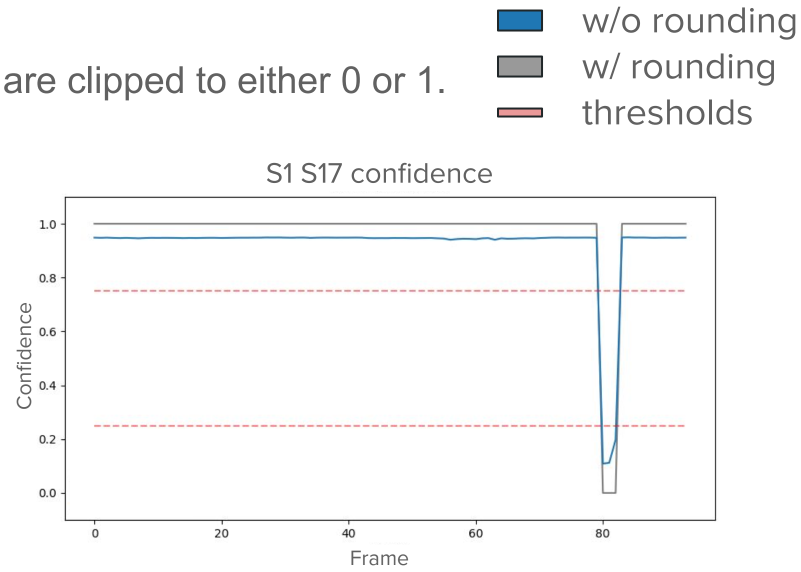
- Segmentation
  - Use RITnet_v3 pre-trained model of EllSeg on OpenEDS, NVGaze, RITEyes, LPW, Fuhl and PupilNet.
  - Fine-tune RITnet_v3 on S1-S4 dataset.
- Classification
  - Use MAE pre-trained model on ImageNet-1k.
  - Random augmentation
    - Perspective warping
    - Brightness
    - Horizontal flipping



Method — post-processing

- Area testing: using segmentation map to correct "marginal" tasks.
    - The variation in pupil area within the same folder is not significant.
- Clipping and rounding
    - The confidence values passing the thresholds are clipped to either 0 or 1.

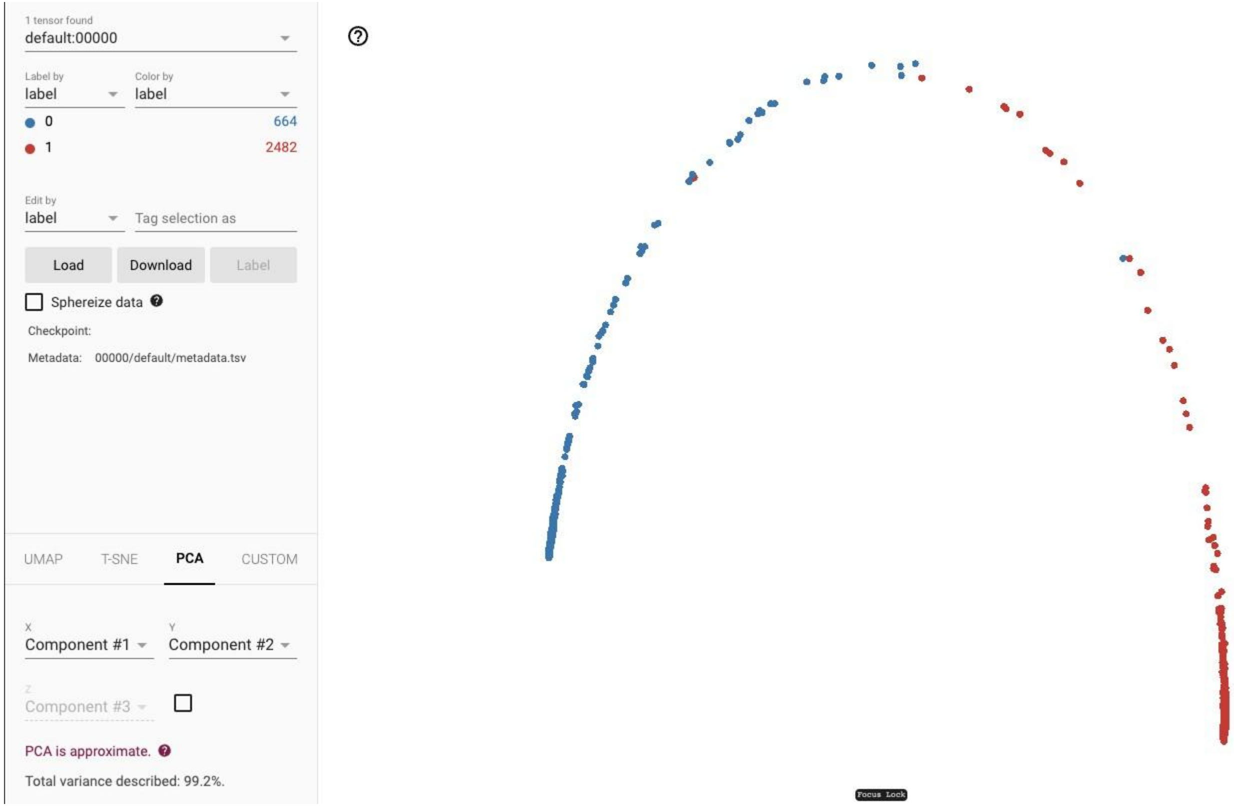




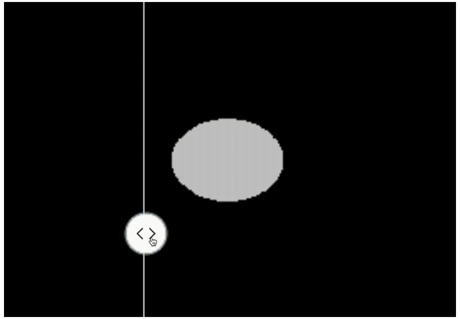
Experiment

- Ellipse fitting video visualization and 數據
- MAE
  - t-SNE結果圖
  - augmentation
    - 亮度
    - 透視旋轉
    - 旋轉
    - 對比度
    - randomcrop
- RIT
  - fine-tune前後的比較 (0.66->0.67)
- Adding visualization of video of segmentation and confidence

Experiment


1 tensor found
default:00000
Label by
label
Color by
label
0
664
1
2482
Edit by
label
Tag selection as
Load
Download
Label
Sphereize data
Checkpoint:
Metadata: 00000/default/metadata.tsv
UMAP
T-SNE
PCA
CUSTOM
X
Component #1
Y
Component #2
Z
Component #3
PCA is approximate.
Total variance described: 99.2%.
Focus Lock

Experiment — Segmentation



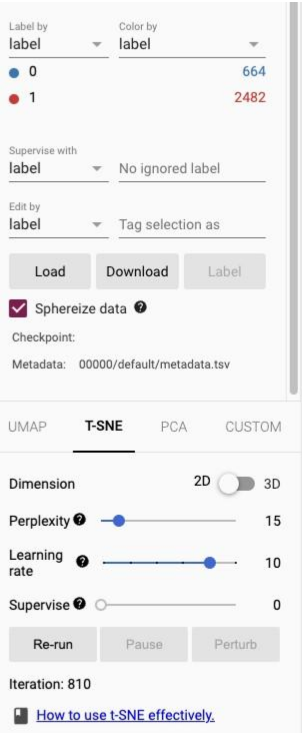
- To improve the performance in terms of mIoU,
    a. Fine-tune on the competition dataset on the base of the pretrained model.
    b. Replace the original output with the smooth ellipse.

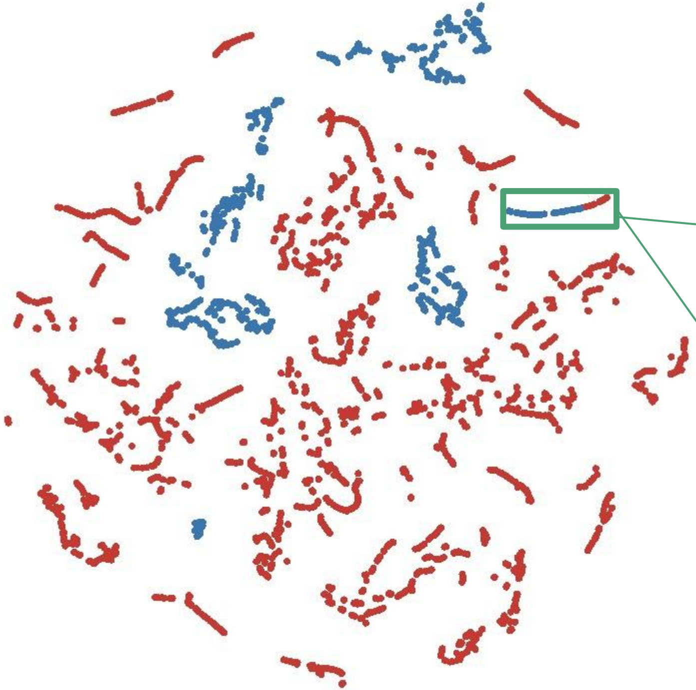
- The comparisons on the testing dataset.
    - mIoU increases 1.2% with fine-tuning.
    - Score increases 0.3% with ellipse fitting.

	w/o fine-tuning	w/ fine-tuning
mIoU	0.664	**0.676**

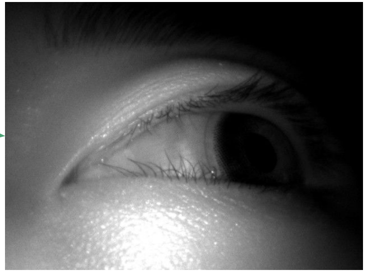
	w/o ellipse fitting	w/ ellipse fitting
score	0.929	**0.932**

Experiment — Classification

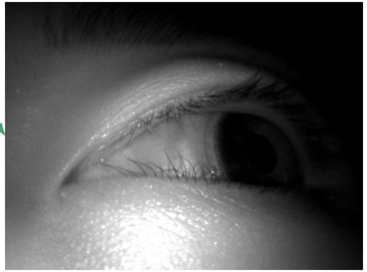




label 0: S4/01/81.jpg



label 1: S4/01/80.jpg



The t-SNE visualization of features extracted by MAE-trained ViT-L on the given S4 dataset

Data augmentation

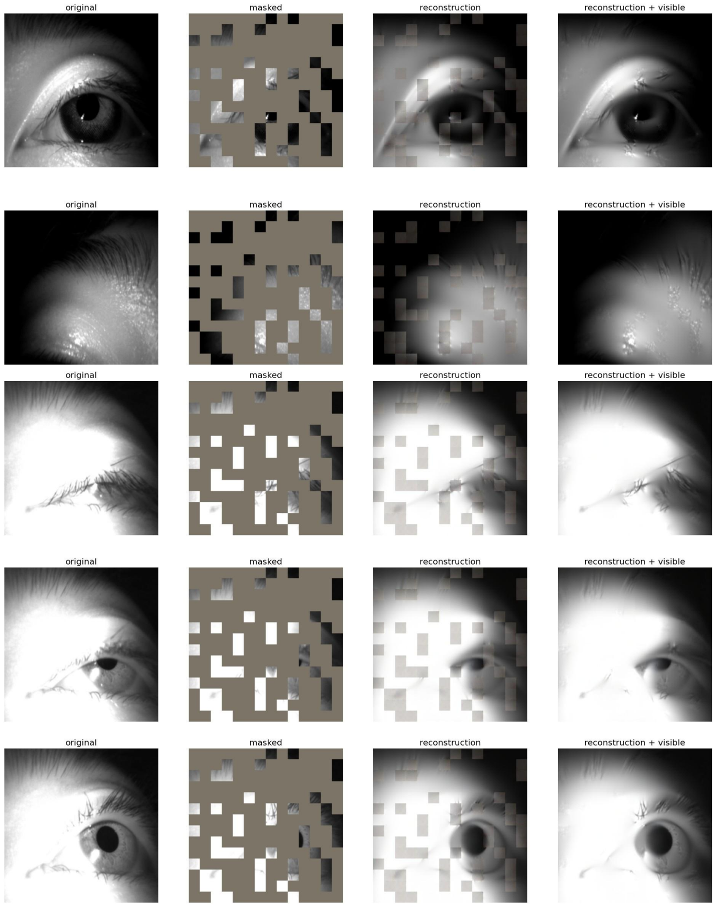
	Performance	Reason
Brightness	⬆	Randomly change the brightness improves robustness in the dataset due to its varying levels.
Contrast	⬇	Unauthorized modification of image contrast may expose previously undetected content to the model.
Rotation	⬇	Increasing the diversity of the data can be achieved without affected the ground truth on confidence
Perspective	⬆	Increasing the diversity of the data can be achieved without change the existence of the pupil
RandomCrop	⬇	There is a possibility that it may disrupt the label.

Buffer

S4/04/0210 開

S4/25/0157 閉

S1/02/0142~0144 開～閉



Buffer

Conclusion

- With RITnet_v3 pretrained weights and fine-tuning, we achieved an mIoU of 0.965 on the testing set.
- With pretrained weights from MAE and sensible augmentation techniques, we achieved 0.950 on the leaderboard.
- By applying suitable post-processing, we can detect some corner cases of the pupil existence based on the raw segmentation map.

Q&A

Thanks for your attention.

Future work

- To utilize the robust features learned by MAE, SegViT (NIPS 22) can be used to perform the segmentation task.
- Leveraging the shared ViT structure enables multi-task learning, allowing a single model to simultaneously perform pupil segmentation and occlusion classification.