Semi-Supervised Learning for Eye Image Segmentation

Aayush K.Chaudhary* (akc5959@rit.edu), Prashnna K. Gyawali* (pkg2182@rit.edu),
Linwei Wang, Jeff B. Pelz
Rochester Institute of Technology, NY, USA

RIT Rochester Institute of Technology
College of Science Chester F. Carlson Center for Imaging Science
Golisano College of Computing and Information Sciences
ETRA 2021
Computational Biomedicine Lab
Multidisciplinary Vision Research Laboratory

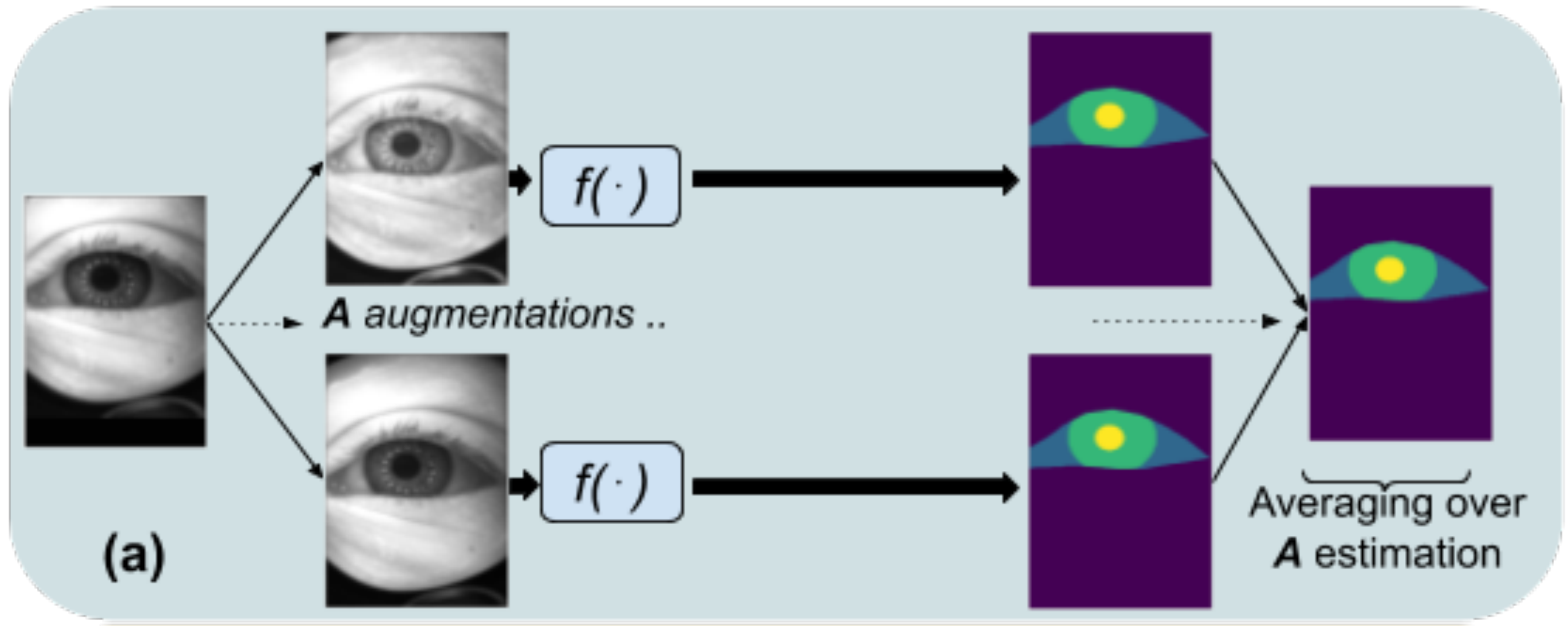
Overview

- Labeling **large** dataset is **tedious,** requires expertise [leads to] bias and inconsistency
- Semi-supervised learning (SSL) : Exploit **hidden relationships** within data to predict labels for **unlabeled** images
- Network **Consistency : Same label prediction** to image even after **perturbation** (mostly used in **classification** task)
- For **semantic segmentation** assumption of network consistency **is violated**

Proposed Method

SSL with domain-specific augmentation (SSL$_D$)

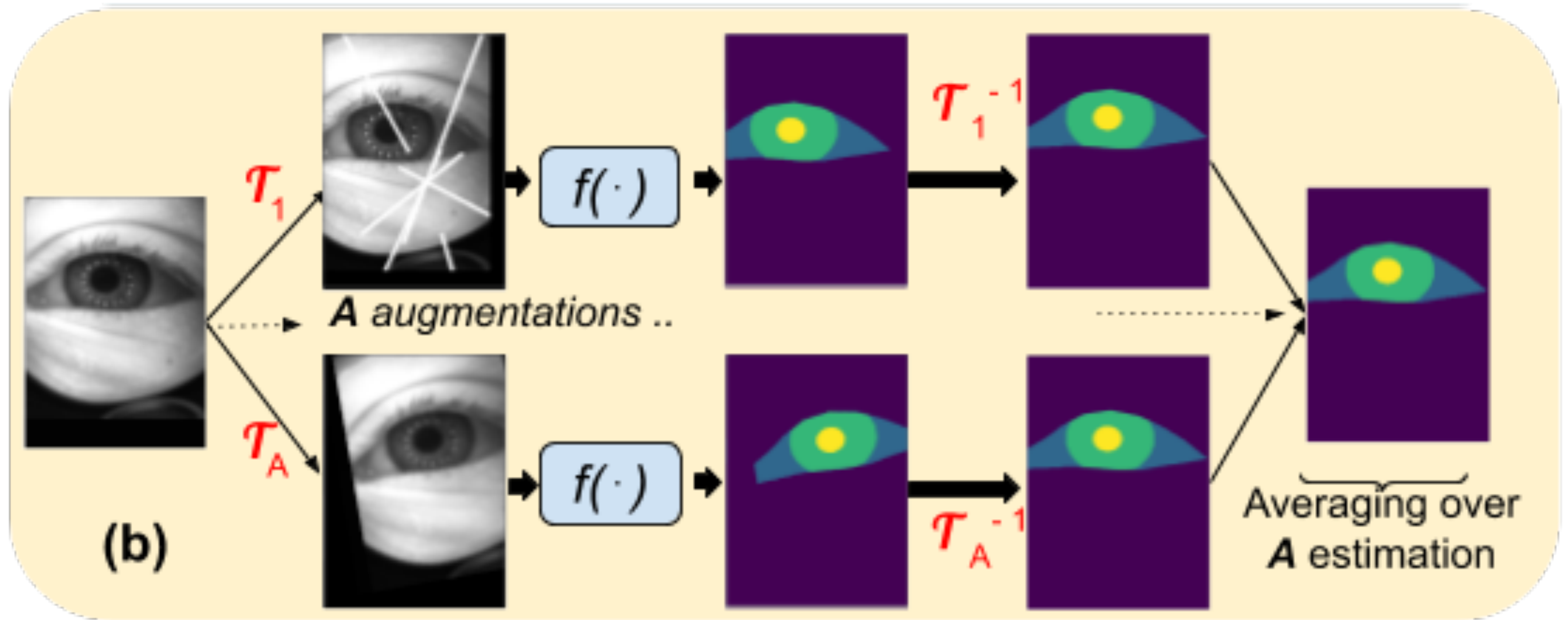
- Vary **contrast & luminance** of eye images
- **CLAHE & Gamma correction**



For each unlabeled image, labels are guessed for A separate copies generated via proposed SSL with domain-specific augmentation

SSL with self-supervised Learning (SSL$_{SS}$)

- **Pretext learning task** -> predicting image from inversion of the transformed image
- Account for **translation / rotation** of images



For each unlabeled image, labels are guessed for A separate copies generated via proposed SSL with a self-supervised approach

Objective function

Total loss = Supervised loss +
$\lambda u \times$ **unsupervised loss +**
$\lambda ss \times$ **self-supervised loss**

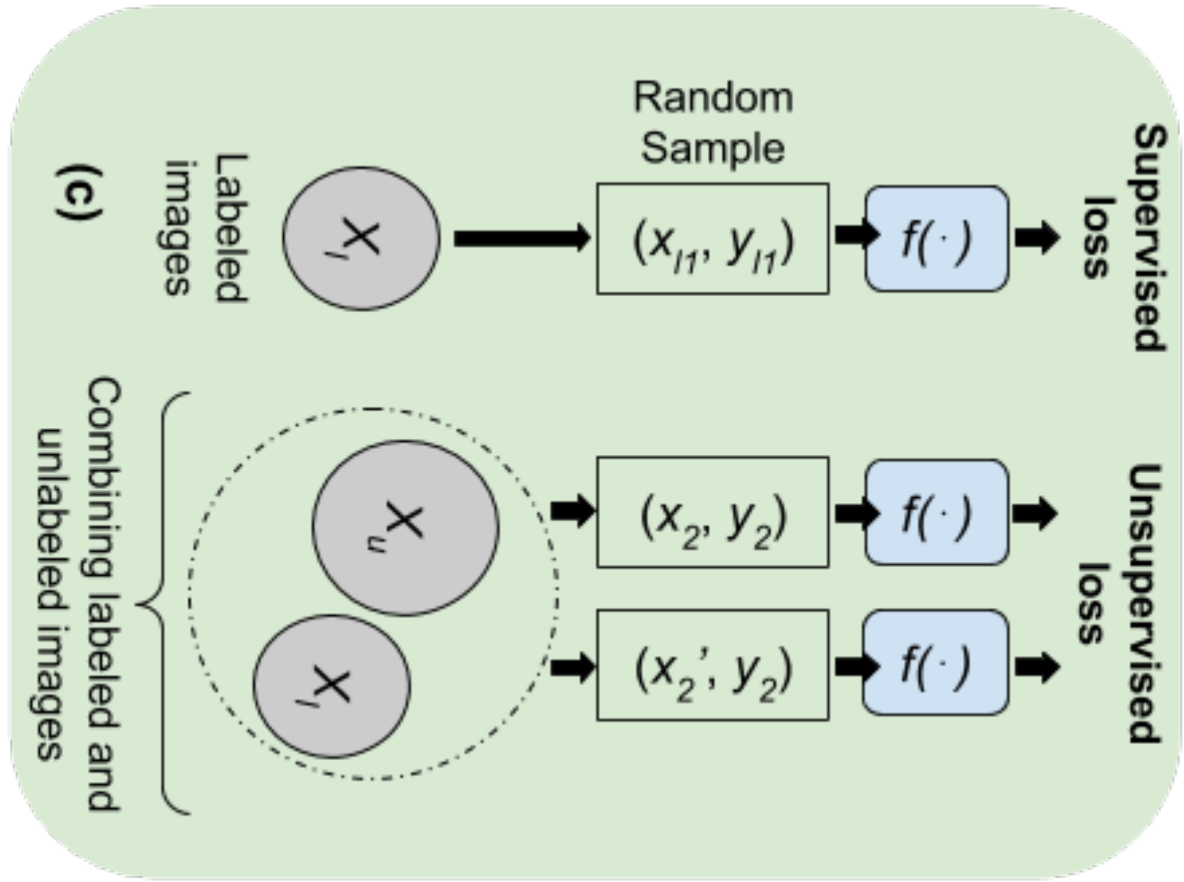


Fig: Supervised loss and unsupervised loss are computed separately for labeled and a combination of labeled and unlabeled data set in both types of SSL methods.

Results

Training with multiple subjects

- fixed unlabeled images (8916 images)
- for X_l= 8916 images, S_L achieved 94.80% whereas we obtained **94.73% with SSL$_{SS}$ for X_l = 940 images only.**

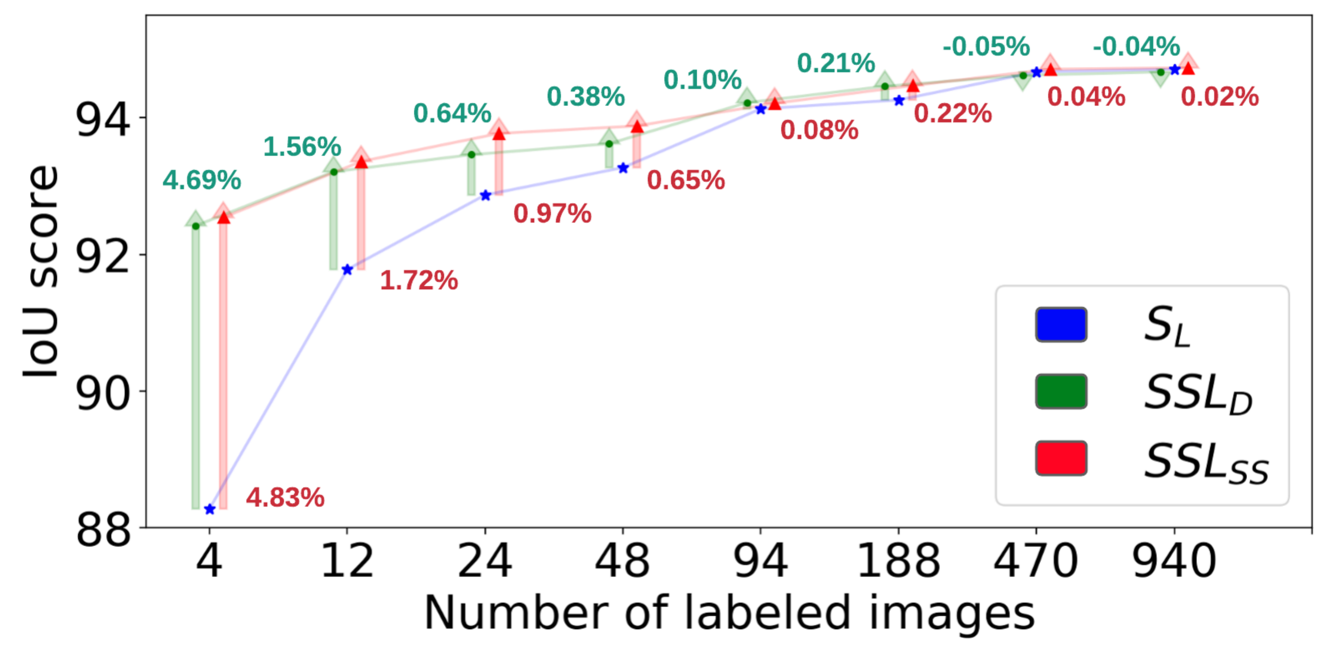


Fig: IoU score for three cases (S_L: blue, SSL_D: green, and SSL_{SS}: red) are shown with varying numbers of X_l and fixed X_u. The number alongside arrows indicate respective improvement (in %) over S_L.

Our Presence at ETRA 2021

- Enhancing the precision of remote eye-tracking using iris velocity estimation (Short Paper)
- Privacy-Preserving Eye Videos using Rubber Sheet Model (Short Paper)
- RIT-Eyes, realistically rendered eye images for eye tracking applications (Video)

Training with single subject

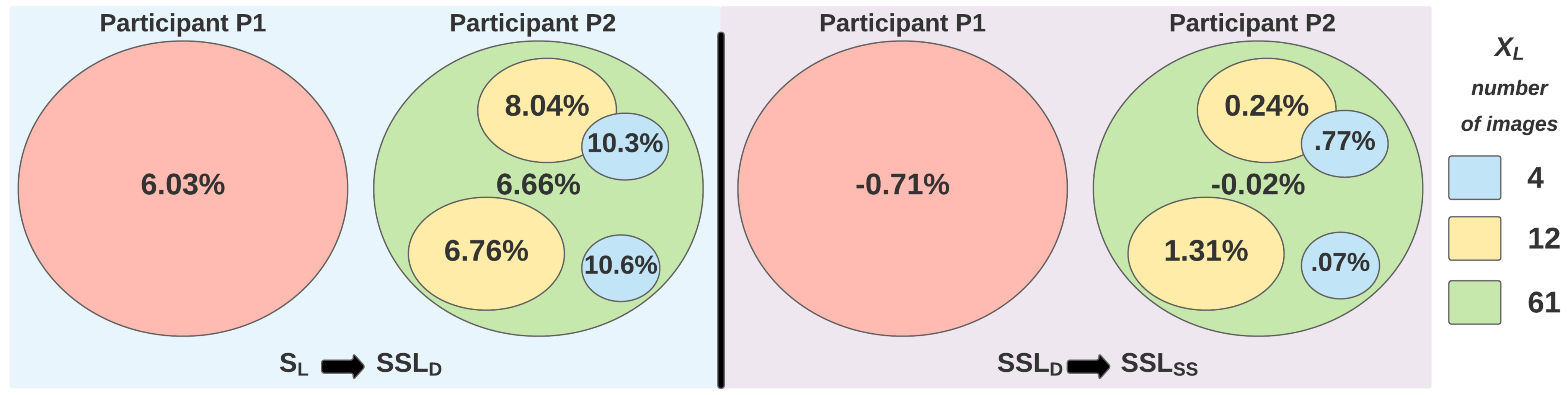


Fig: Demonstration of improvement (in %) for cases S_L to SSL_D (left) and SSL_D to SSL_{SS} (right) when models are trained on two subjects (red and green). For P2 (green), we further test the change in model performance for various subsets of X_l .

Eye part Segmentation

X_l	4	12	24	48	4 (P1)	12 (P1)	61 (P1)
% change	4.08 (4.48)	0.85 (1.37)	0.54 (0.99)	0.44 (0.78)	4.40 (7.32)	4.50 (6.12)	2.13 (4.09)

Fig: Comparison of % change of pupil and iris (inside parenthesis) class IoU scores for cases from S_L to SSL_{SS} for varying number of X_l and fixed number of X_u. P1 indicates samples from a single subject.

Qualitative Results

- **As the number of images increases**, the confidence in prediction and unwanted spurious patches are reduced when models are trained on S_L.
- For SSL approaches, **the confidence in prediction is more even when a small number of X_l are used.**
- **No significant difference** is visible for the **two SSL approaches**, which vary mostly in **finer details.**

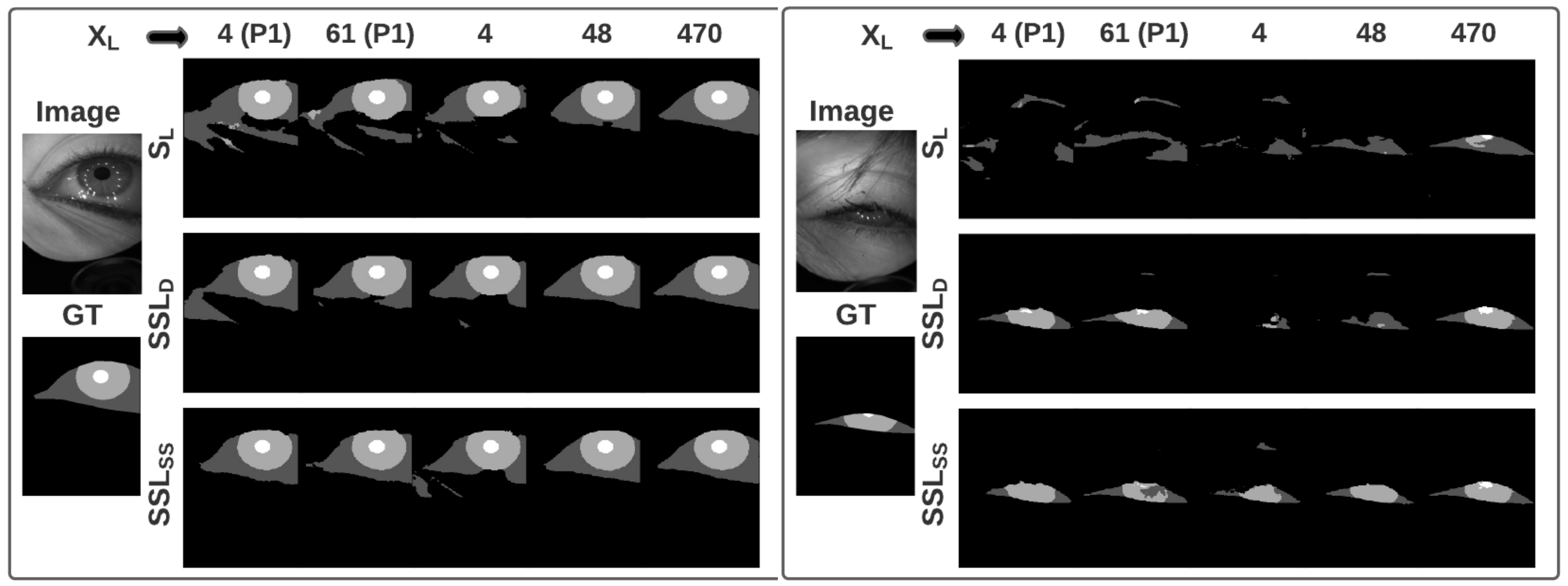


Fig: Two samples of the test set with its corresponding ground-truth and network predictions for the number of cases are shown in adjacent blocks.

Conclusion

- Frameworks to leverage **domain specific** augmentations and **novel spatially varying image transformations**
- Trained on just 4 and 48 labeled images, improvement by at least 4.7% and 0.4% respectively, in segmentation performance
- Future Work - Investigate the effect of curating labeled datasets (e.g., **considering eye position and blinks**) instead of random selection