Reverse Knowledge Distillation: Training a Large Model using a Small One for Retinal Image Matching on Limited Data

Sahar Almahfouz Nasser*, Nihar Gupte*, and Amit Sethi

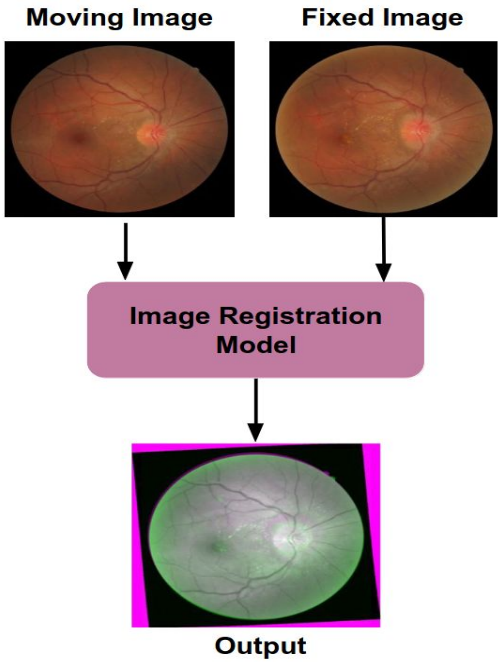
Electrical Engineering Department, Indian Institute of Technology Bombay

* Indicates equal contribution

Introduction

- **Retinal Image Matching (RIM)**
  - plays a crucial role in monitoring disease progression and treatment response

- **RIM is challenging due to:**
  - Variations in blood vessels, optic nerve position, and other features
  - Pathological Changes
  - Limited overlap
  - Non-rigid transformation
  - Inter-subject variabilities
  - Real-time processing: demanding efficient and rapid algorithms

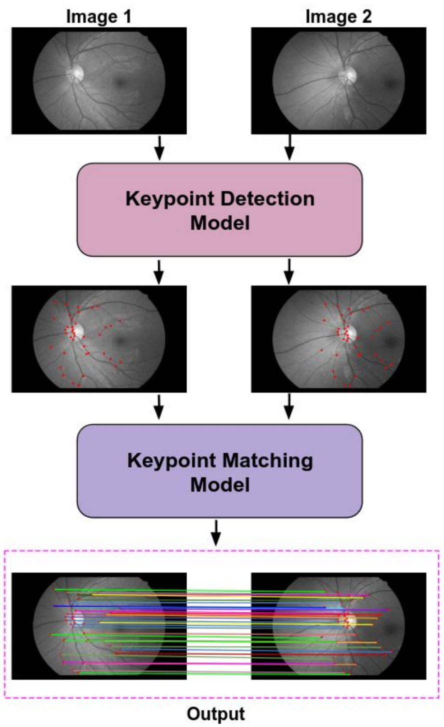


(Source: Sabanovic et al, 2017)[1]

[1] Sabanovic, E. and Matuzevicius, D., 2017, November. Experimental investigation of feature descriptors for retinal image registration. In *2017 5th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)* (pp. 1-4). IEEE.

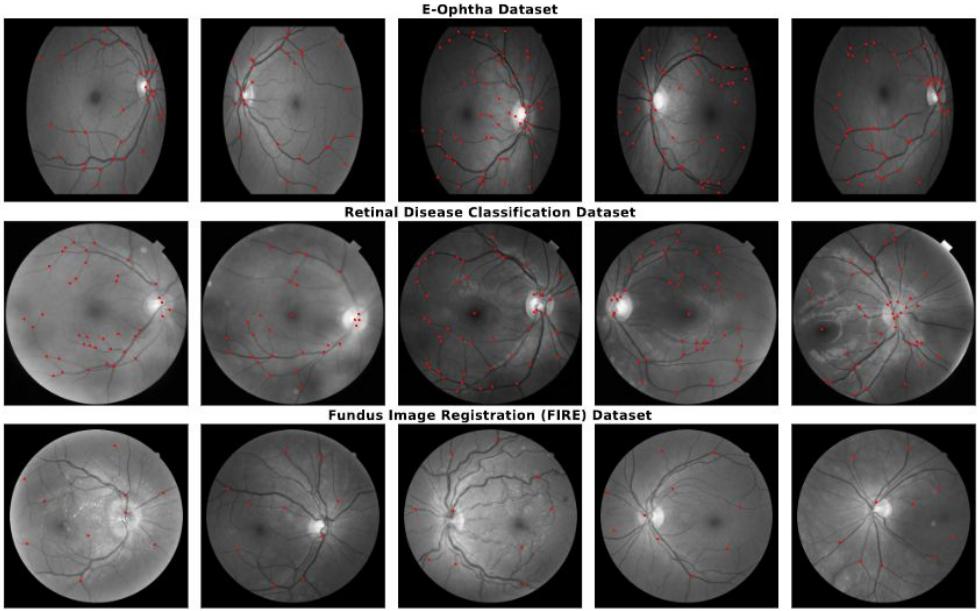
RIM Pipeline

- **Keypoint detection and feature extraction**
- **Traditional keypoint detection methods**
  - Harris corner detector, SIFT, SURF
  - Drawbacks:
    - Time consumption
    - Limited accuracy under lighting and viewpoint changes, occlusions and cluttered backgrounds
- **DL-based keypoint detectors**
  - Oriented fast and rotated BRIEF (ORB)
  - SuperPoint
  - Low dimensional step pattern analysis (LoSAP)
  - Greedily Learned Accurate Match Points (GLAMpoints)
  - SuperRetina (SOTA)



Datasets

- **MeDAL-Retina dataset** [1]
  - 261 normal images (Train/Val: 208/61)
  - Annotations: intersections, crossovers, and bifurcations
  - Avg. number of keypoints: 42.96 ± 14.03
  - Sources: 201 from e-ophtha, 60 images from retinal disease classification dataset
  - 1.9K images collected from public resources
  - Data Preparation: z-score normalization, CLAHE, Gamma correction
- **FIRE dataset for testing only** [2]
  - 129 images of three categories: S, P, A
  - S: 71 pairs, overlap>75%, minimal anatomical differences
  - P: 49 pairs, significant differences (shift, rotation)
  - A: 14 pairs, images acquired at different examinations



A Visual comparison between MeDAL-Retina [1] and FIRE [2] datasets

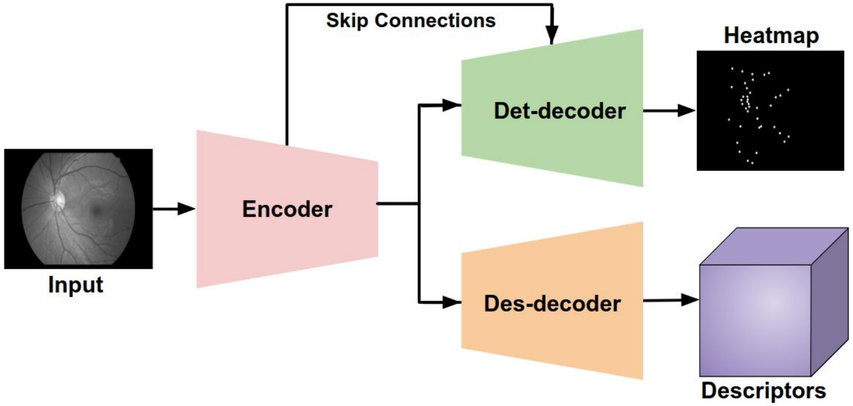
Thanks to Nihar, Prateek, Keshav, Tanmay for helping in dataset Preparation

[1] Gupte, N., Almahfouz Nasser, S., Garg, P., Singhal, K., Jain, T., Aditya, Kumar, R., & Sethi, A. (2023). MeDAL-Retina [Dataset]. Retrieved from [https://www.dropbox.com/sh/o8q84e2eg54ay3d/AADiAkNr6bFQDoFaKeEjpYtra?dl=0]

[2] Carlos Hernandez-Matas, Xenophon Zabulis, Areti Triantafyllou, Panagiota Anyfanti, Stella Douma, and Antonis A Argyros. Fire: fundus image registration dataset. Modeling and Artificial Intelligence in Ophthalmology, 1(4):16–28, 2017.

Proposed Method

- SuperRetina[1] **Semi-supervised** learning
- **Architecture:** encoder, keypoint detector, keypoint descriptor
- **Types:**
  - Unet-empowered SuperRetina
  - Large kernel-empowered SuperRetina (Ours1)
  - Swin UNETR-empowered SuperRetina (Ours2)



The general architecture of SuperRetina

[1] Jiazhen Liu, Xirong Li, Qijie Wei, Jie Xu, and Dayong Ding. Semi-supervised keypoint detector and descriptor for retinal image matching. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXI, pages 593–609. Springer, 2022

Proposed Method: UNet-based SuperRetina

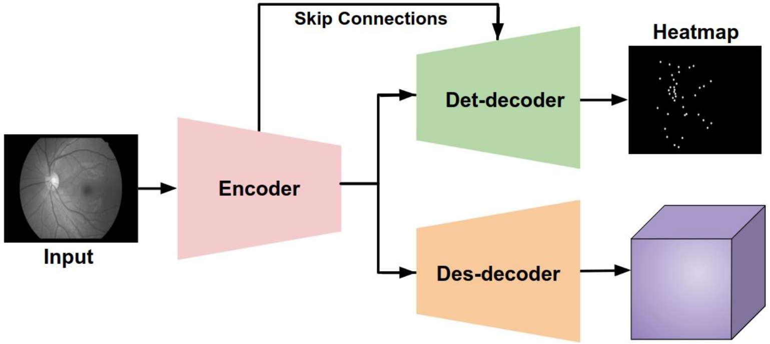
$$l_{total} = l_{det} + l_{des}$$

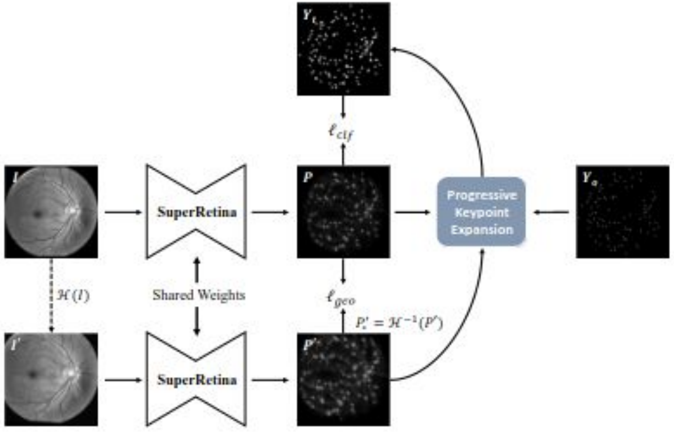
$$l_{det} = l_{clf} + l_{geo}$$

$$l_{clf}(I;Y) = 1 - \frac{2.\sum_{i,j}(P \circ \tilde{Y})_{i,j}}{\sum_{i,j}(P \circ P)_{i,j} + \sum_{i,j}(\tilde{Y} \circ \tilde{Y})_{i,j}}$$

$$l_{des}(I,H) = \sum_{(i,j)\in\tilde{P}} max(0, m + \Phi_{i,j} - \frac{1}{2}(\Phi_{i,j}^{rand} + \Phi_{i,j}^{hard}))$$

l_{clf}: Dice-based classification loss
l_{geo}: Dice-based geometric loss
l_{des}: Descriptors loss
$\tilde{Y}$: Smoothed version of the binary ground truth labels Y
P: Keypoint heatmap
Φ: Distance value



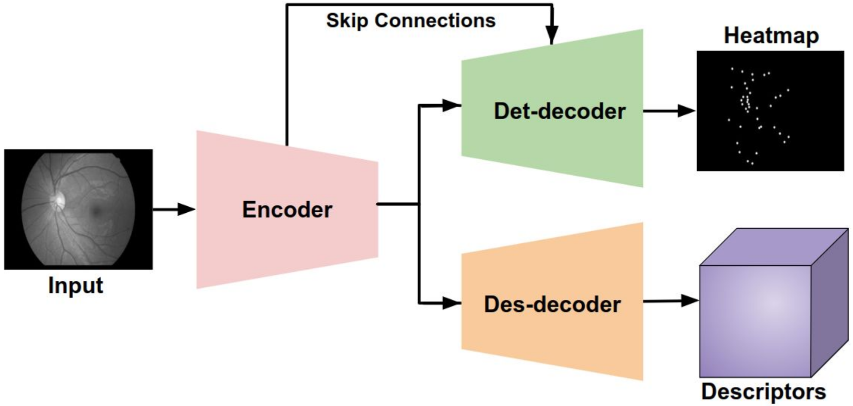


Geometric Loss: Credits [1]

[1] Jiazhen Liu, Xirong Li, Qijie Wei, Jie Xu, and Dayong Ding. Semi-supervised keypoint detector and descriptor for retinal image matching. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXI, pages 593–609. Springer, 2022

Proposed Method: Large Kernel-based SuperRetina

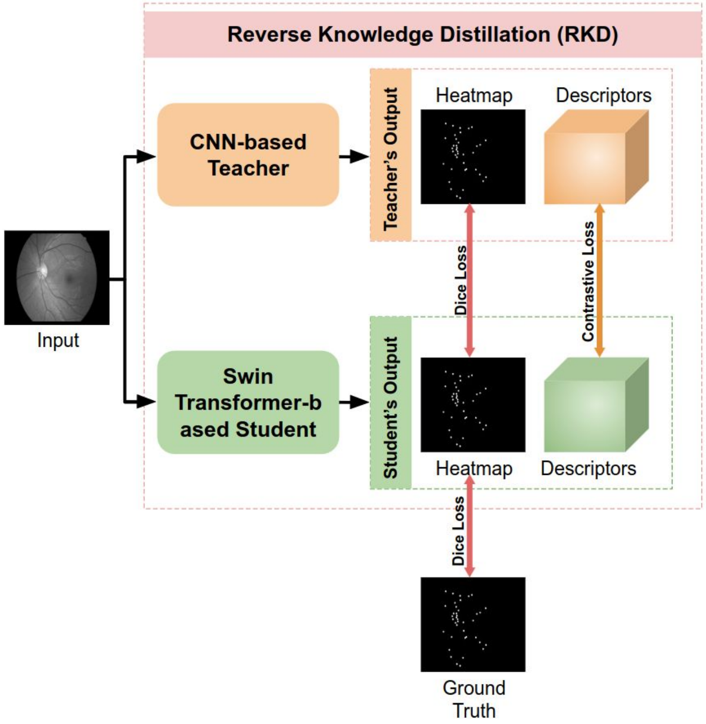
- **Large kernel-empowered SuperRetina**
  - Introducing kernels of various sizes in each of the encoder's layers
  - Capturing long range dependencies
  - Kernels: 1x1, 3x3, 5x5
  - SOTA in terms of mAUC



The general architecture of SuperRetina

Proposed Method: Transformer-based SuperRetina

- **Swin UNETR-empowered SuperRetina**
  - A transformer-based encoder
  - Swin transformer and CNN in a Unet-style architecture
  - Reverse Knowledge distillation
    - A teacher (CNN) guides a student (transformer)
    - Generalization: drop out 50%



Loss Function

$$l_{det} = l'_{clf} + l_{geo} \quad (1)$$

$$l'_{clf} = l_{clf} + l^{RKD}_{clf} \quad (2)$$

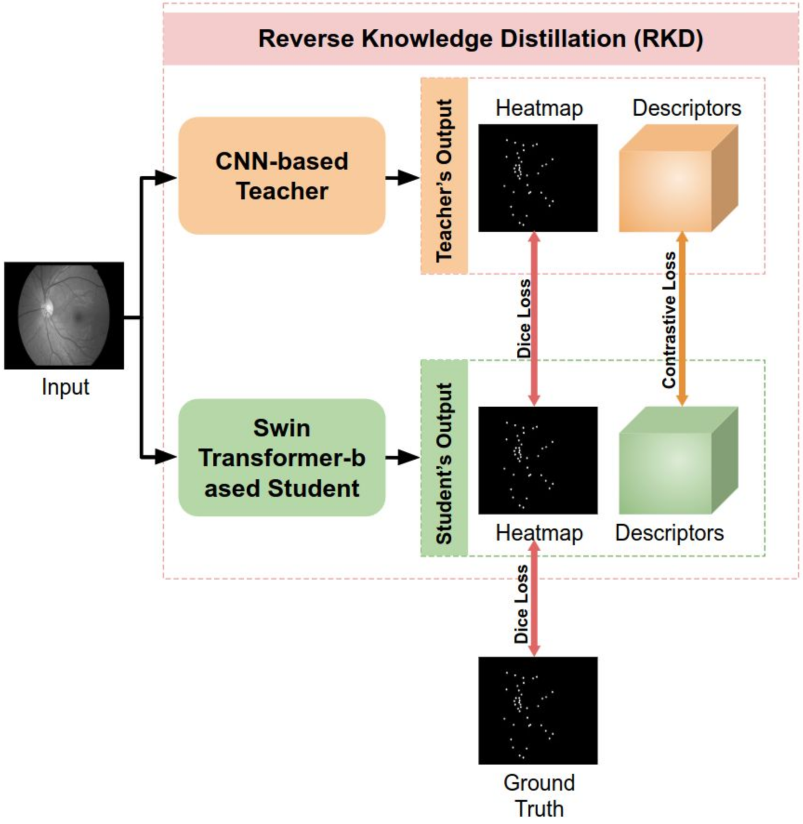
$$l_{clf}(I; Y) = 1 - \frac{2.\sum_{i,j}(P \circ \tilde{Y})_{i,j}}{\sum_{i,j}(P \circ P)_{i,j} + \sum_{i,j}(\tilde{Y} \circ \tilde{Y})_{i,j}} \quad (3)$$

$$l^{RKD}_{clf}(I_S; I_T) = 1 - \frac{2.\sum_{i,j}(P_S \circ P_T)_{i,j}}{\sum_{i,j}(P_S \circ P_S)_{i,j} + \sum_{i,j}(P_T \circ P_T)_{i,j}} \quad (4)$$

$$l_{Des} = l_{des} + l^{RKD}_{des} \quad (5)$$

$$l_{des}(I, H) = \sum_{(i,j)\in\tilde{P}} max(0, m + \Phi_{i,j} - \frac{1}{2}(\Phi^{rand}_{i,j} + \Phi^{hard}_{i,j})) \quad (6)$$

l_{clf}: Dice-based classification loss
l_{geo}: Dice-based geometric loss
l_{des}: Descriptors loss
$\tilde{Y}$: Smoothed version of the binary ground truth labels Y
P_S: Keypoint heatmap of the student
P_T: Keypoint heatmap of the teacher model
$\tilde{P}$: Non-maximum supressed keypoint set for each keypoint (i, j)
Φ: Distance value



Evaluation Metrics

- **Failure rate**
- **Acceptance rate**
- **The median distance**
- **The maximum distance**
- **AUC (easy, moderate, hard, and mean)**

Results

Method	Failed	Inaccurate	Acceptable	AUC-Easy	AUC-Mod	AUC-Hard	mAUC
SIFT, IJCV04 [25]	**0.00%**	20.15%	79.85%	0.903	0.474	0.341	0.573
PBO, ICIP10 [26]	0.75%	28.36%	70.89%	0.844	0.691	0.122	0.552
REMPE, JBHI20 [18]	**0.00%**	02.99%	97.01%	**0.958**	0.660	0.542	0.720
SuperPoint, CVPRW18 [13]	**0.00%**	05.22%	94.78%	0.882	0.649	0.490	0.674
GLAMpoints, ICCV19 [37]	**0.00%**	07.46%	92.54%	0.850	0.543	0.474	0.622
R2D2, NIPS19 [28]	**0.00%**	12.69%	87.31%	0.900	0.517	0.386	0.601
SuperGlue, CVPR20 [34]	0.75%	03.73%	95.52%	0.885	0.689	0.488	0.687
NCNet, TPAMI22 [29]	**0.00%**	37.31%	62.69%	0.588	0.386	0.077	0.350
SuperRetina [23]	**0.00%**	01.50%	98.50%	0.940	**0.783**	0.542	0.755
Ours-1 (Large kernel-SuperRetina)	**0.00%**	00.75%	99.25%	0.942	**0.783**	**0.558**	**0.761**
Ours-2 (Swin UNETR-SuperRetina)	**0.00%**	**00.00%**	**100.0%**	0.935	0.780	0.550	0.755

The superior method is determined by having a higher acceptance rate or AUC, and lower rates of inaccuracies or failures

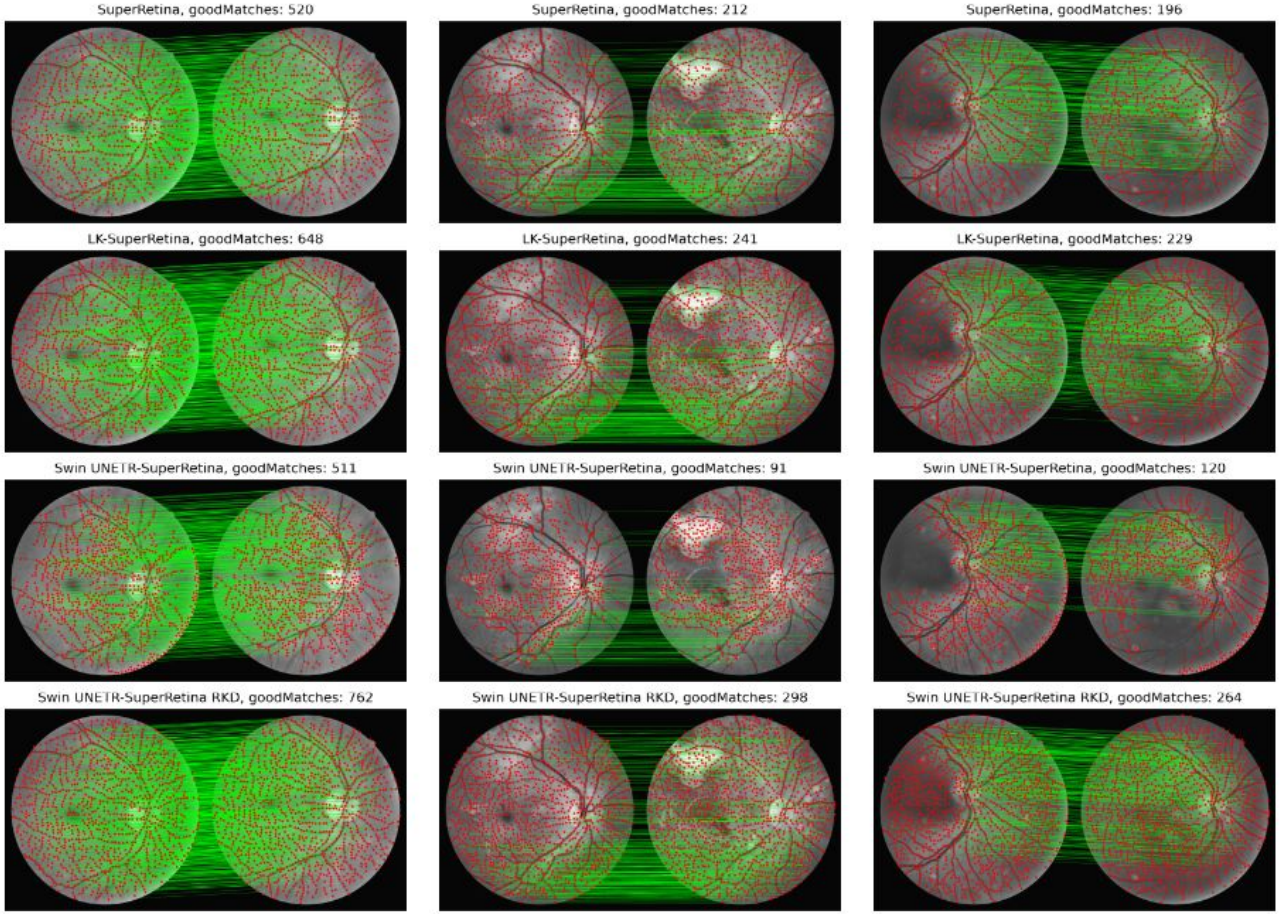
Ablation Studies

Method	Failed	Inaccurate	Acceptable	AUC-Easy	AUC-Mod	AUC-Hard	mAUC
SuperRetina [22], KS 3×3	**0.00%**	01.50%	98.50%	0.940	**0.783**	0.542	0.755
LK-SuperRetina, KS $1 \times 1, 3 \times 3, 5 \times 5$	**0.00%**	00.75%	99.25%	0.942	**0.783**	0.558	**0.761**
LK-SuperRetina, KS $1 \times 1, 3 \times 3, 5 \times 5, 7 \times 7$	**0.00%**	02.25%	97.74%	0.925	0.717	0.502	0.714
Swin UNETR-SuperRetina, Trained from scratch	**0.00%**	16.55%	83.45%	0.891	0.649	0.318	0.619
Swin UNETR-SuperRetina, SuperRetina as teacher w/o dropout (DO)	**0.00%**	01.5%	98.50%	**0.947**	0.769	0.549	0.755
Swin UNETR-SuperRetina, SuperRetina as teacher, DO 50%	**0.00%**	**00.00%**	**100.0%**	0.935	0.780	0.550	0.755
Swin UNETR-SuperRetina, LK-SuperRetina as teacher, DO 50%	**0.00%**	00.75%	99.25%	0.914	0.774	0.558	0.749
Pretrained Swin UNETR-SuperRet., LK-SuperRet. as teacher, DO 50%	**0.00%**	00.75%	99.25%	0.928	0.774	**0.559**	0.754

- 50% dropout, resulting in a significant performance boost for the Swin UNETR-empowered SuperRetina
- **100%** accuracy on the testing dataset
- RKD model has **2.5%** accuracy boost over the baseline

Results

- RKD model has more number of good matches for all categories (Easy, Moderate, and Hard)



Easy **Moderate** **Hard**

Geometric Registration: Image Matching

A. Retinal image matching

B. Face Alignment

Face Alignment

- The Wider Facial Landmarks in-the-wild (WFLW) dataset [1]
- 10,000 faces, with 7,500 for training and 2,500 for testing
- 98 annotated landmarks
- Wide range of variations
- Loss is MSE

$$l'_{mse} = l_{mse} + \lambda l^{RKD}_{mse}$$

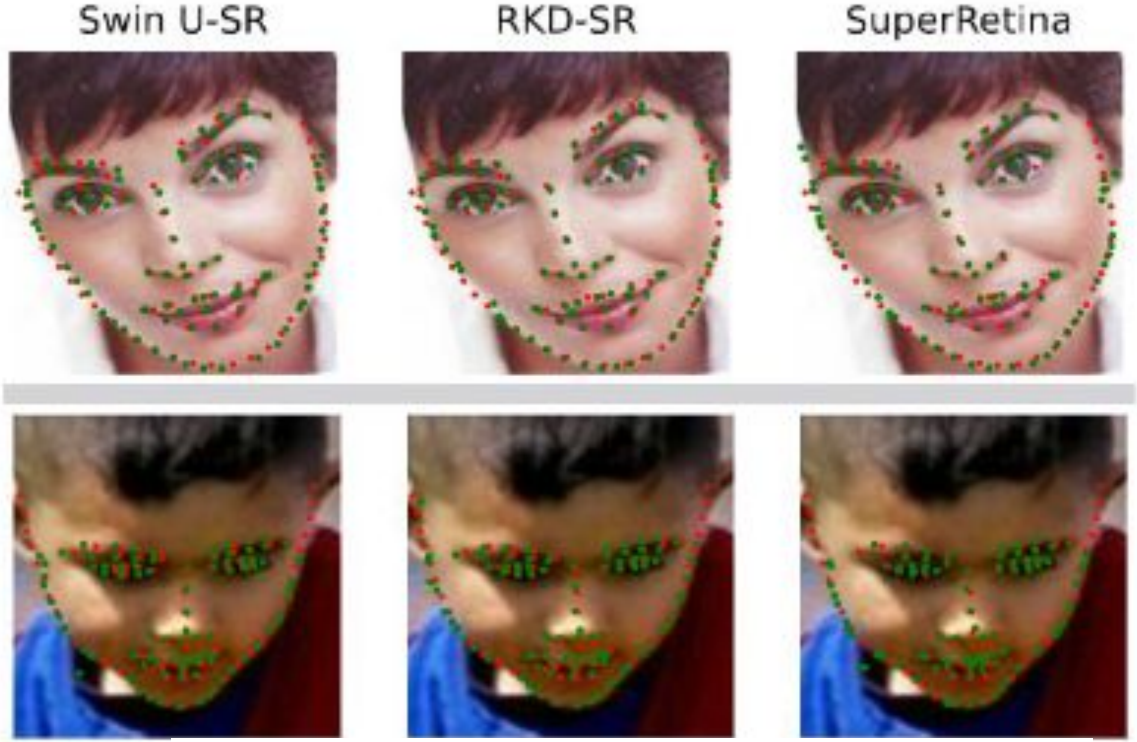
Where Lambda is a balancing factor

Samples from WFLW dataset [1]

[1] Wayne Wu, Chen Qian, Shuo Yang, Quan Wang, Yici Cai, and Qiang Zhou. Look at boundary: A boundary-aware face alignment algorithm. In CVPR, 2018.

Results

- RKD-SR combines the favorable aspects of both models
- Only RKD-SR demonstrates robustness against outliers
- RKD-SR achieves a **9.51%** reduction in normalized mean error (NME) compared to the baseline SuperRetina



Method	SuperRetina	Swin U-SR	RKD-SR
NME(%)	20.43	11.15	**10.92**

Contact Details:
sahar.almahfouz.nasser@gmail.com
www.linkedin.com/in/sahar-almahfouz-nasser

Thank You !