MICCAI2023

Fundus-Enhanced Disease-Aware Distillation Model for Retinal Disease Classification from OCT Images

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

广州市中西医结合医院
Guangzhou Hospital Of Integrated Traditional And West Medicine

WRMT

Lehan Wang, Weihang Dai, Mei Jin, Chubin Ou, and Xiaomeng Li ✉

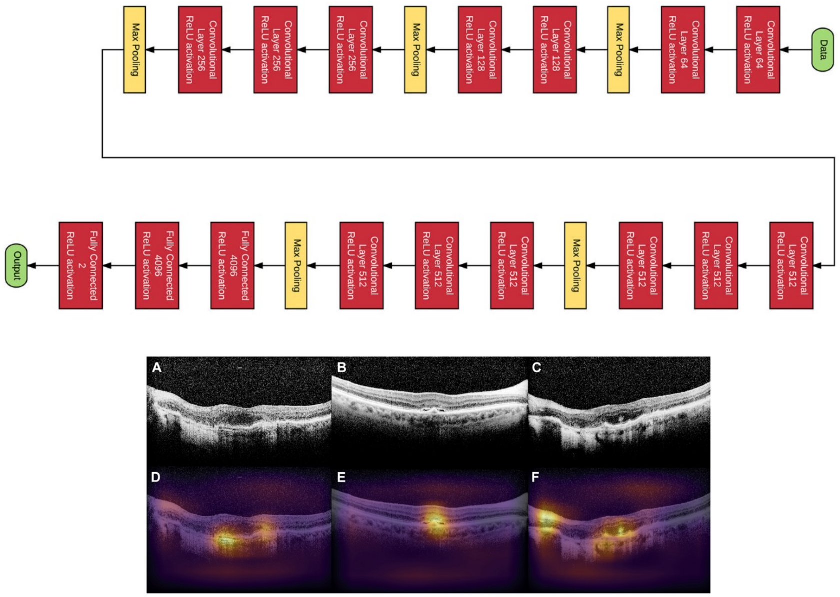
Department of Electronic and Computer Engineering, HKUST

Outline

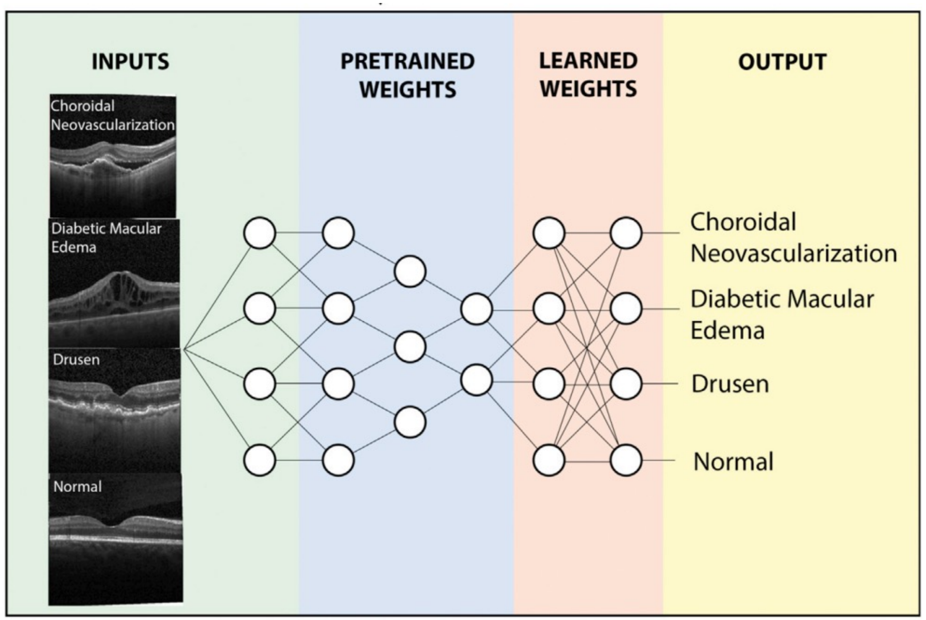
- Introduction
- Methodology
- Experiment
- Future Challenges
- Conclusion

OCT Imaging

- OCT: an innovative imaging technique
  - Used as primary test for many diseases.
  - Also applied to computer-aided diagnosis through deep learning.



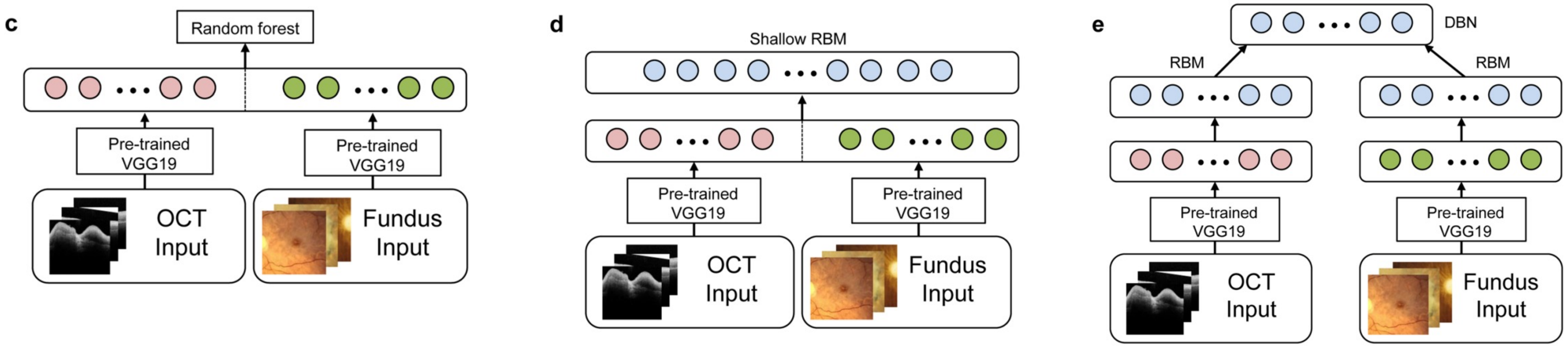
Classify OCT Images into
Normal and AMD
[Lee et al, 2017]



Classification for OCT
into 4 classes
[Kermany et al, 2018]

Multi-Modal Learning

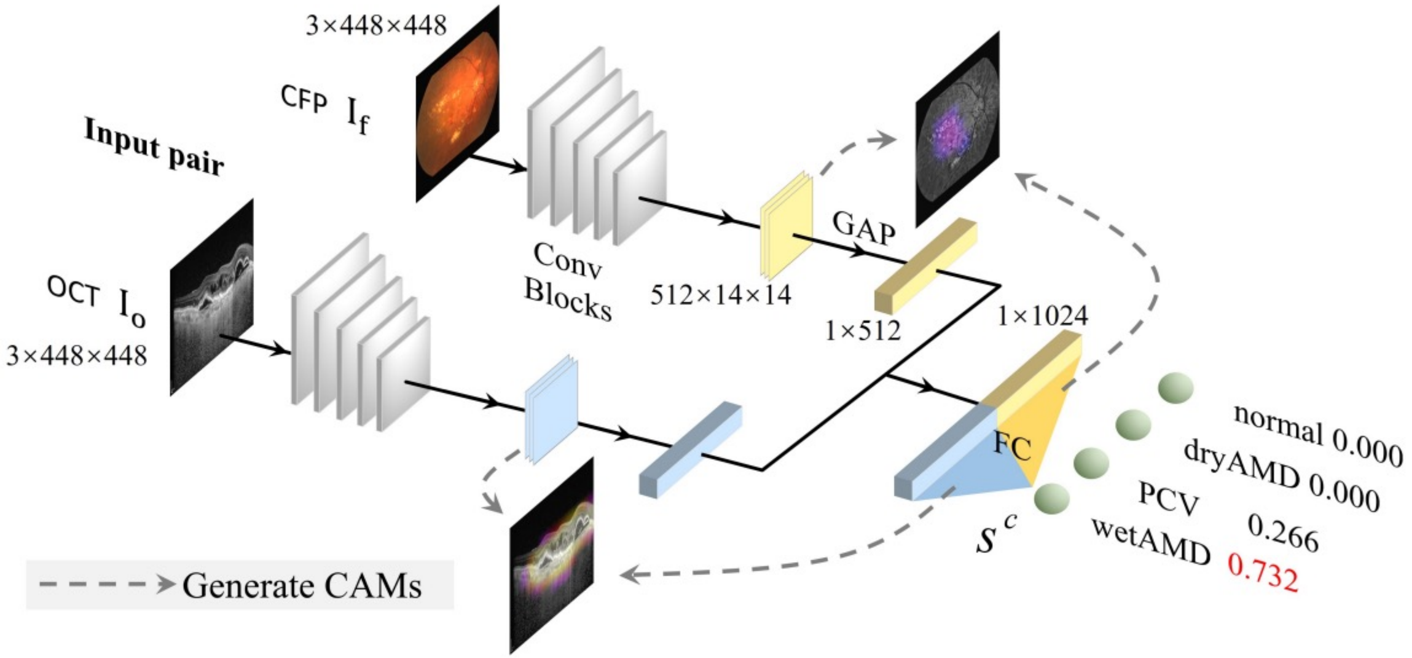
- Multi-modal learning allows to enhance performance of diagnosis
  - Fundus photos are commonly used with OCT images.
  - First attempt to multi-modal learning with OCT and fundus is to detect AMD.



[Yoo et al, 2019]

Related Work

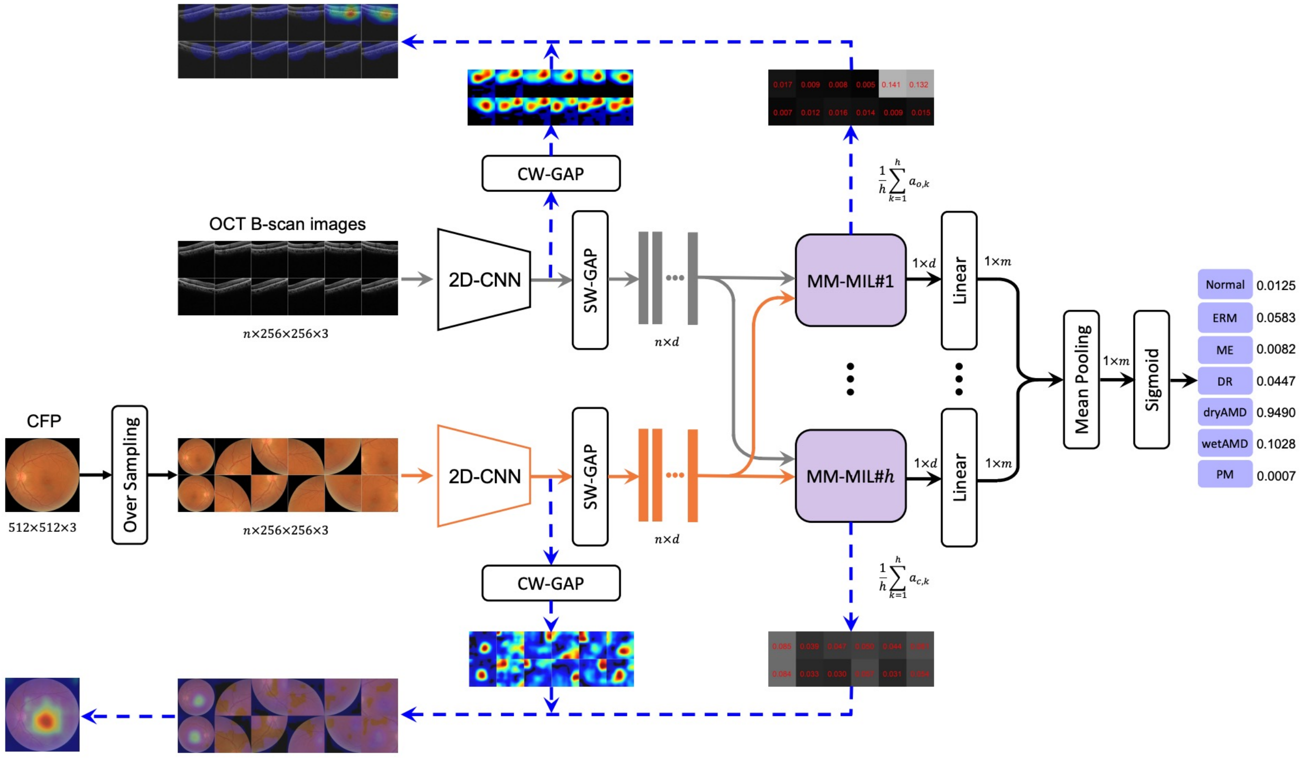
- Two-Stream CNN
  - A two-stream structure to extract fundus and OCT features.
  - Concatenate the features for prediction.

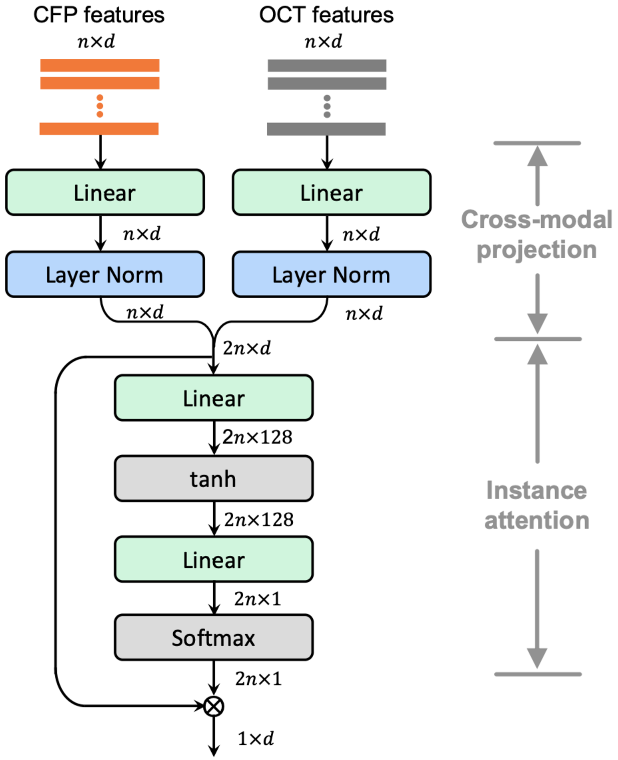


[Wang et al, 2019]

Related Work

- MM-MIL
  - Adopt multi-instance learning to apply all OCT B-scans without selection.
  - Use attention mechanism to fuse OCT and fundus features.

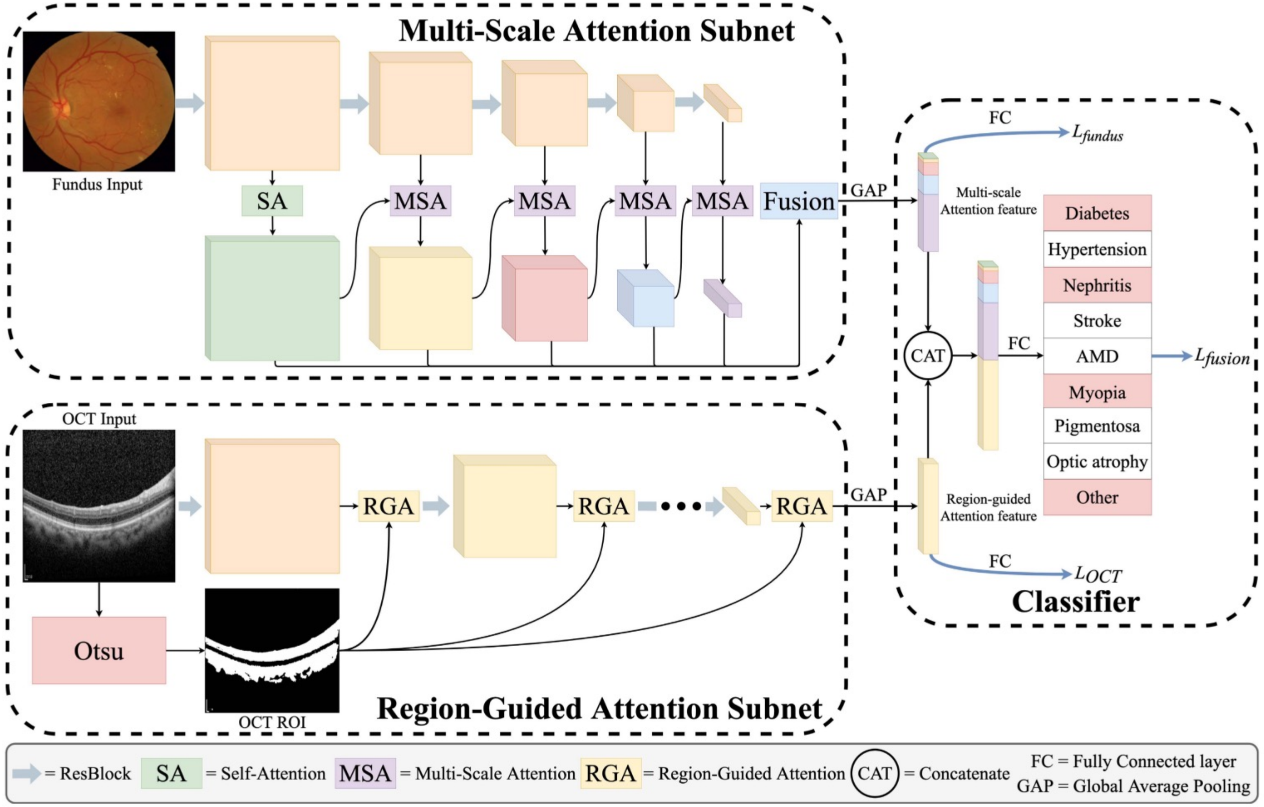




[Li et al, 2021]

Related Work

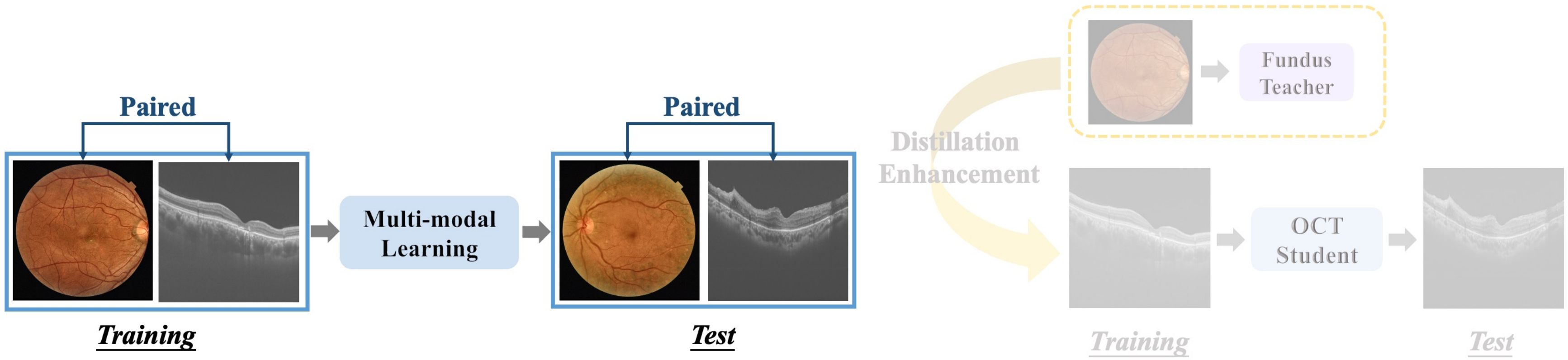
- MSAN
  - Design modality specific network to extract features, considering specific characteristics of different modalities.



[He et al, 2021]

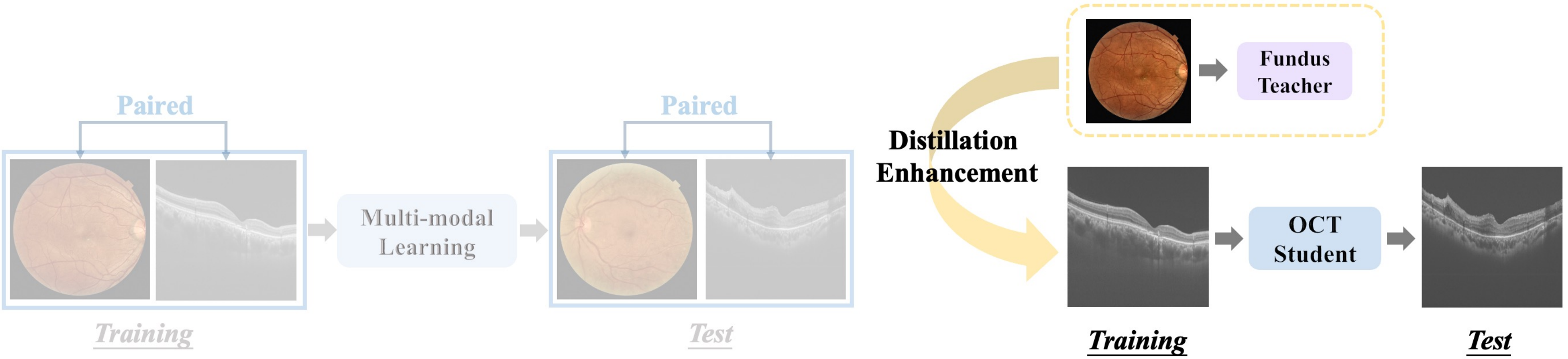
Motivation

- Existing multi-modal learning approaches require strictly paired images from both modalities.
  - The images collected should be paired by patient.
  - This necessitates the collection of multi-modal images for the same patients.
- In previous work, all the modalities are required during both training and test stage.



Contribution

- We propose a novel fundus-enhanced disease-aware distillation model for eye disease classification.
  - Two kinds of distillation are proposed: class prototype matching and class similarity alignment.
- Our approach does not rely on paired training data, and only require OCT images during inference stage.
- Our proposed method offers flexible knowledge transfer from any public available fundus dataset, which can reduce the cost of collecting expensive multi-modal data.

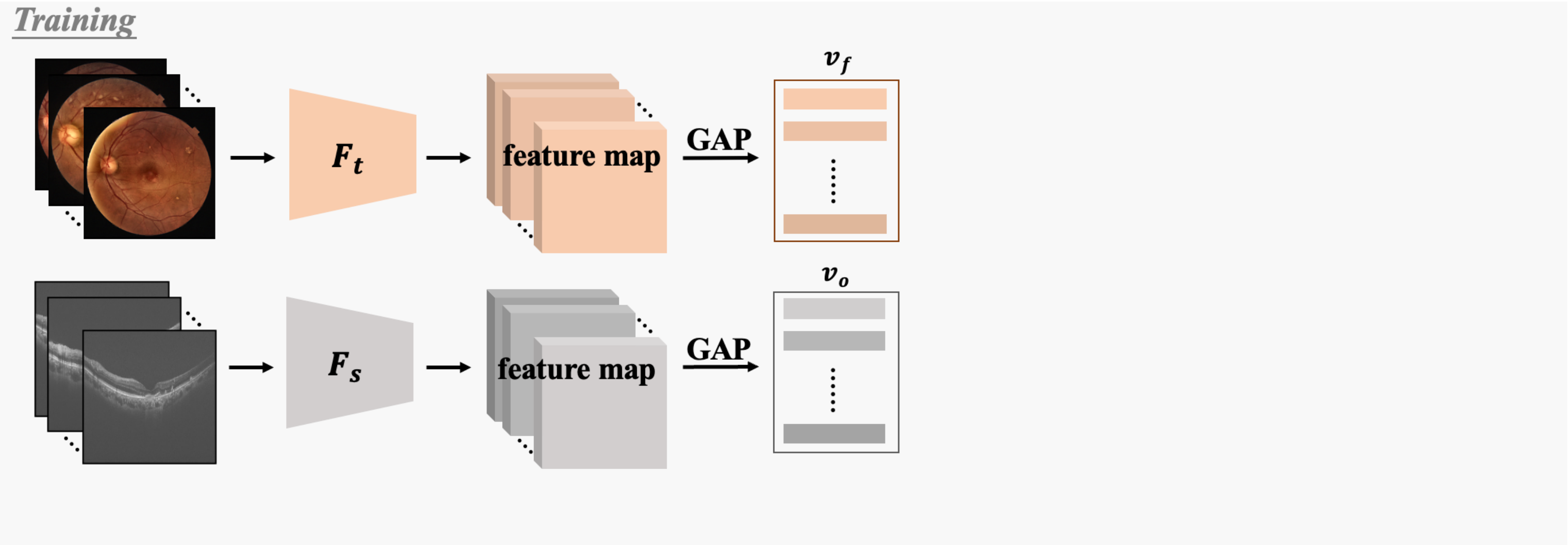


Outline

- Introduction
- Methodology
- Experiment
- Future Challenges
- Conclusion

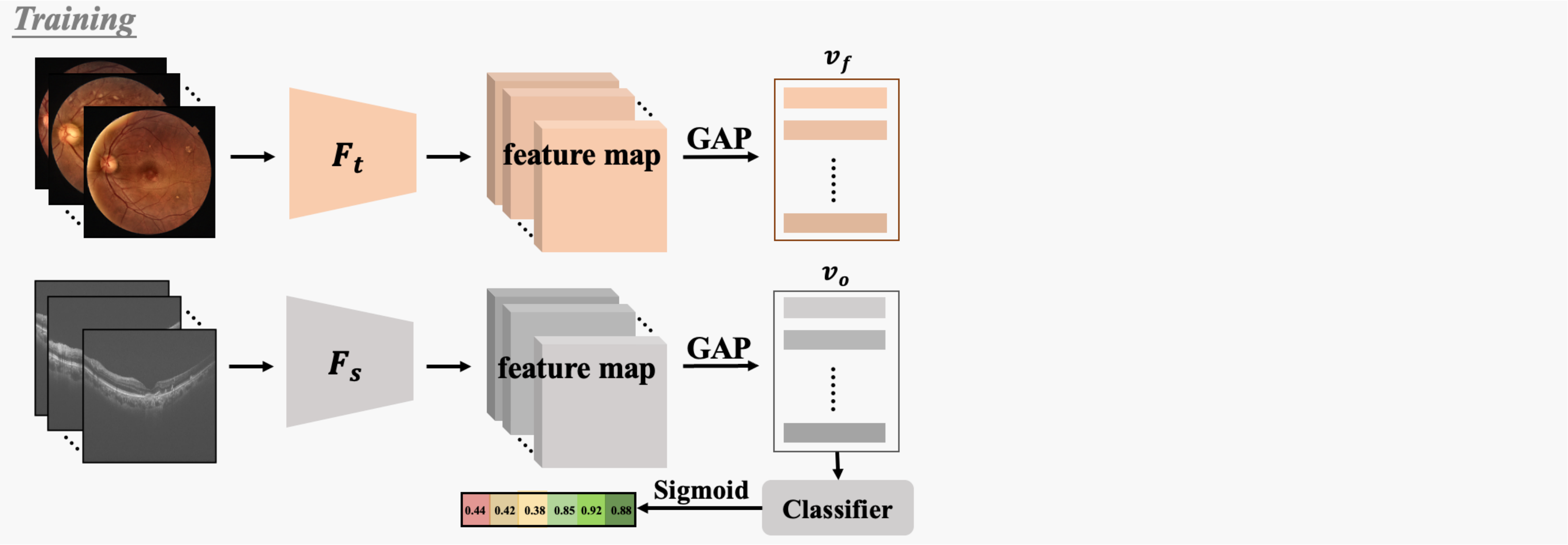
Overview: Framework

- Fundus Feature Encoder & OCT Feature Encoder
  - CNN backbone is used to extract feature vectors.
- Classifier
  - Linear classifier
- Only OCT images are needed in inference stage.



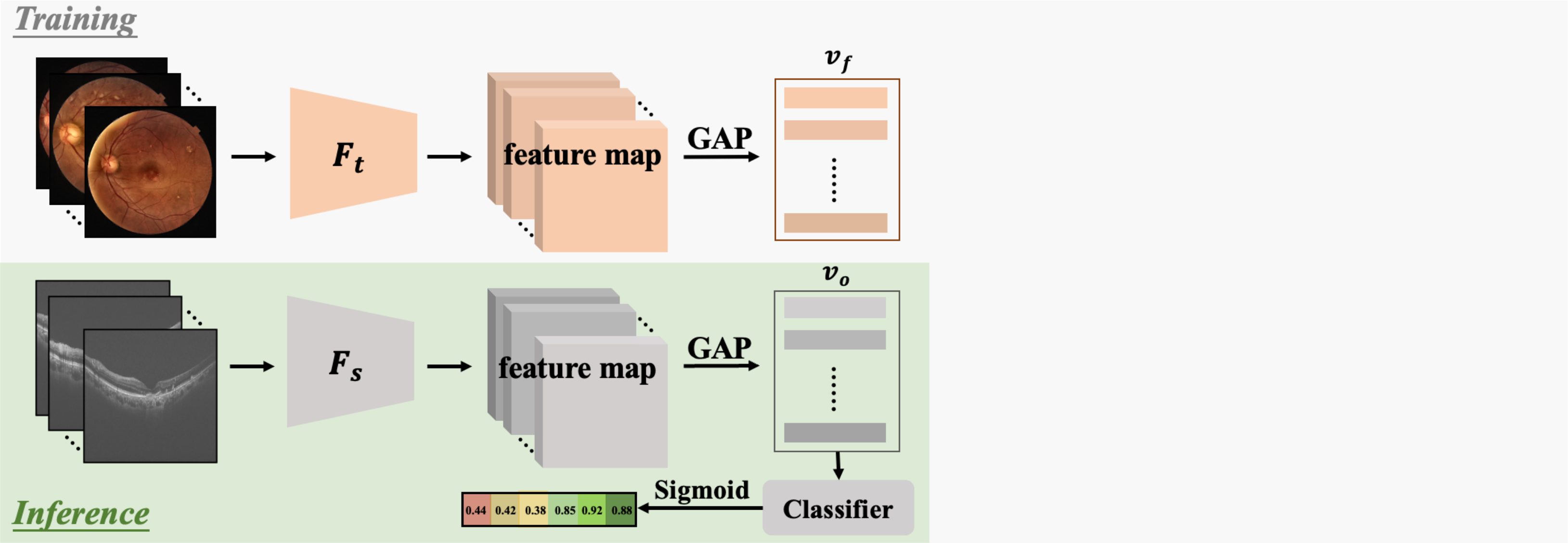
Overview: Framework

- Fundus Feature Encoder & OCT Feature Encoder
  - CNN backbone is used to extract feature vectors.
- Classifier
  - Linear classifier
- Only OCT images are needed in inference stage.



Overview: Framework

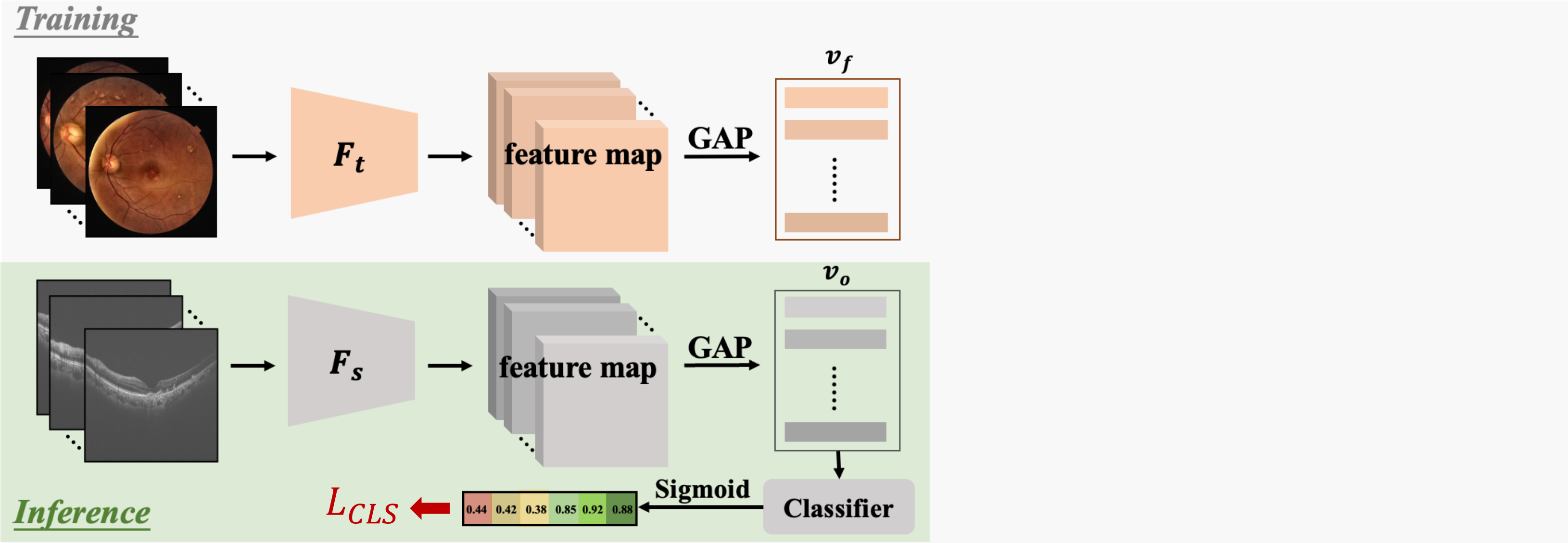
- Fundus Feature Encoder & OCT Feature Encoder
  - CNN backbone is used to extract feature vectors.
- Classifier
  - Linear classifier
- Only OCT images are needed in inference stage.



Overview: Loss Function

$$L_{OCT} = L_{CLS} + L_{DISTILL}$$

- Classification: L_{CLS}
  - Binary Cross Entropy (BCE) is adopted as loss function for classification
- Distillation from fundus model to OCT model: $L_{DISTILL} = \alpha L_{CPM} + \beta L_{CSA}$
  - Class Prototype Matching (CPM): distill disease-related information from fundus model to OCT model.
  - Class Similarity Alignment (CSA): enforce consistency between disease distribution of both modalities.

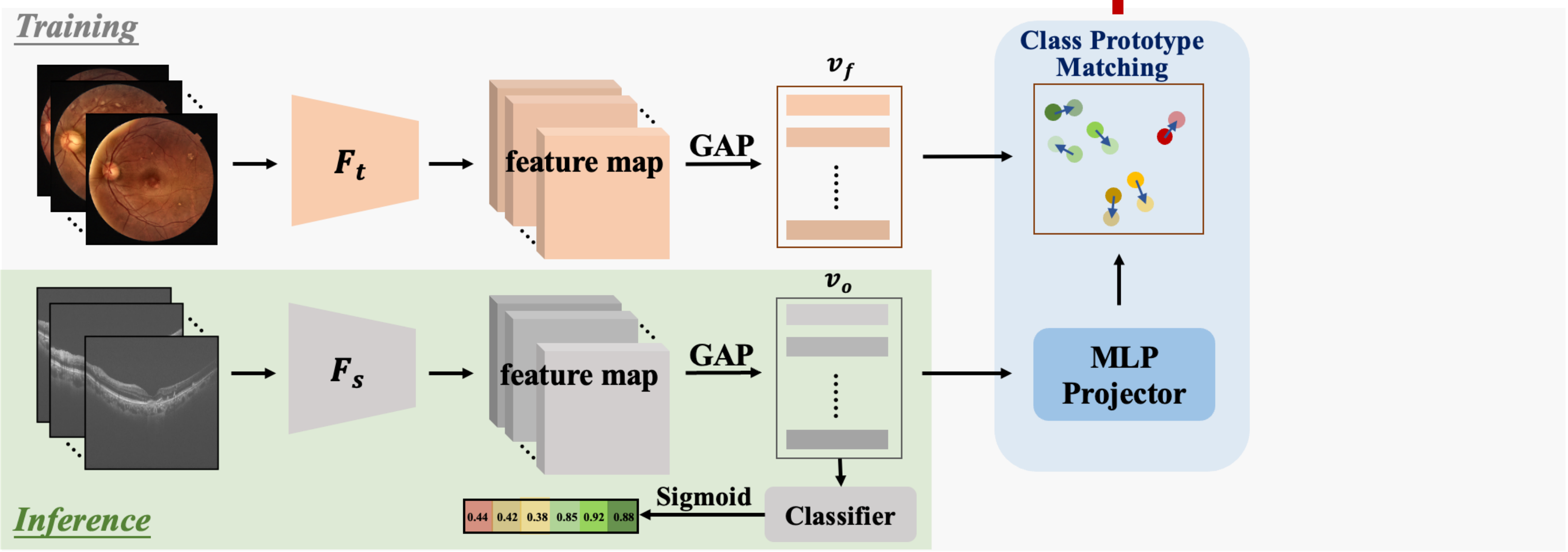


Overview: Loss Function

$$L_{OCT} = L_{CLS} + L_{DISTILL}$$

- Classification: L_{CLS}
  - Binary Cross Entropy (BCE) is adopted as loss function for classification
- Distillation from fundus model to OCT model: $L_{DISTILL} = \alpha L_{CPM} + \beta L_{CSA}$
  - Class Prototype Matching (CPM): distill disease-related information from fundus model to OCT model.
  - Class Similarity Alignment (CSA): enforce consistency between disease distribution of both modalities.

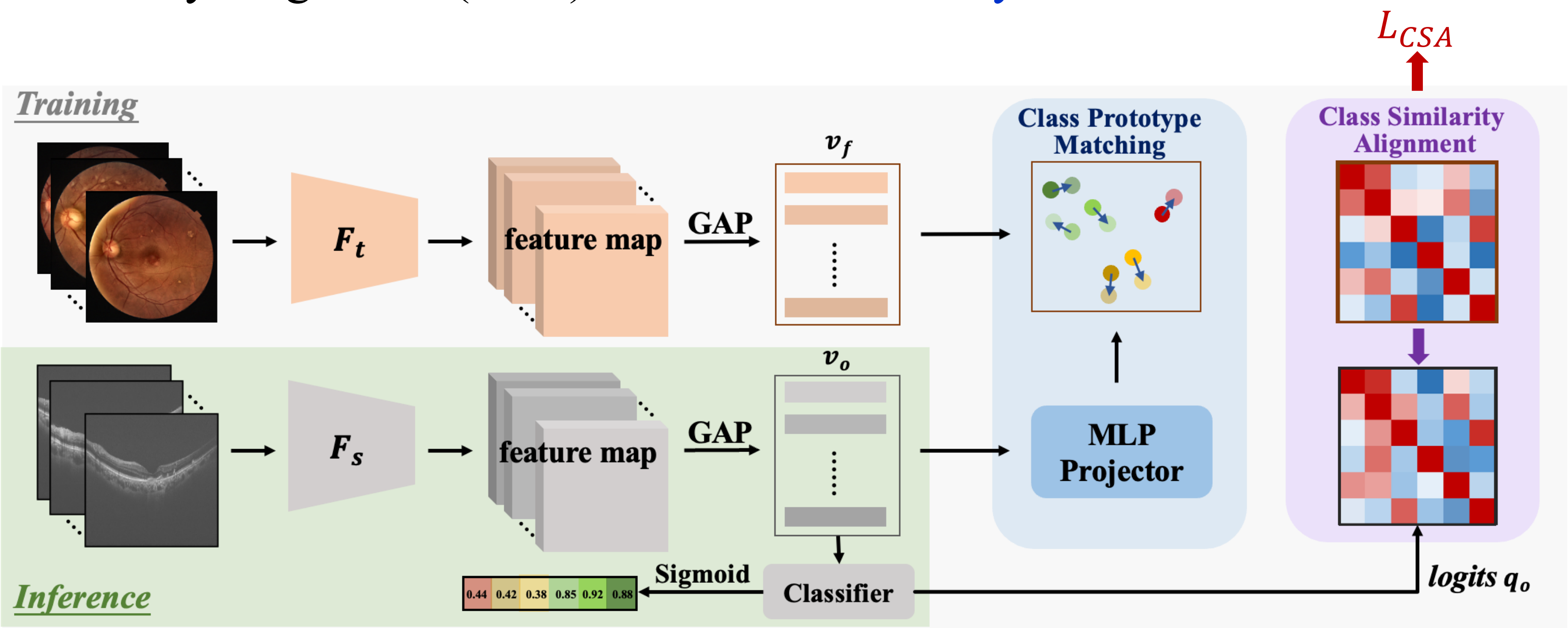
L_{CPM}



Overview: Loss Function

$$L_{OCT} = L_{CLS} + L_{DISTILL}$$

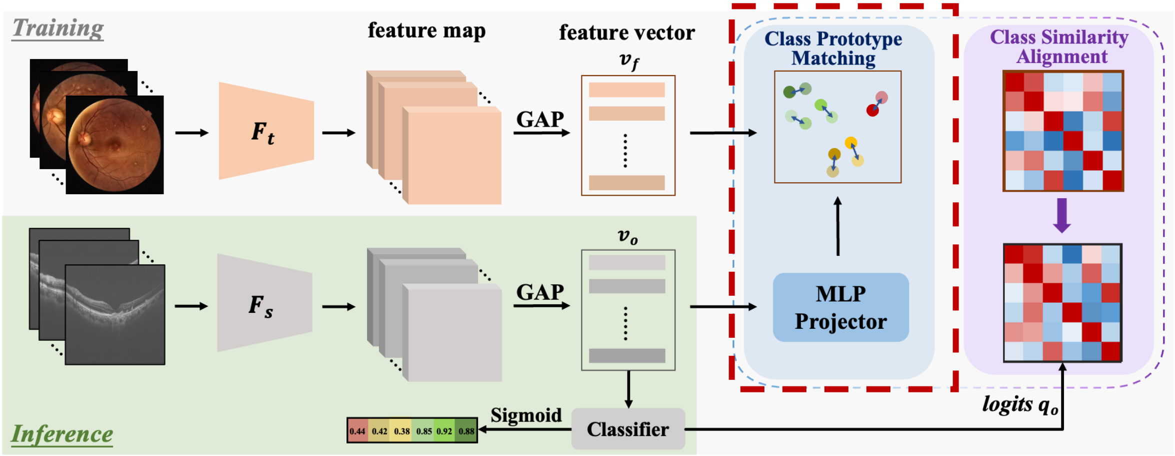
- Classification: L_{CLS}
  - Binary Cross Entropy (BCE) is adopted as loss function for classification
- Distillation from fundus model to OCT model: $L_{DISTILL} = \alpha L_{CPM} + \beta L_{CSA}$
  - Class Prototype Matching (CPM): distill disease-related information from fundus model to OCT model.
  - Class Similarity Alignment (CSA): enforce consistency between disease distribution of both modalities.

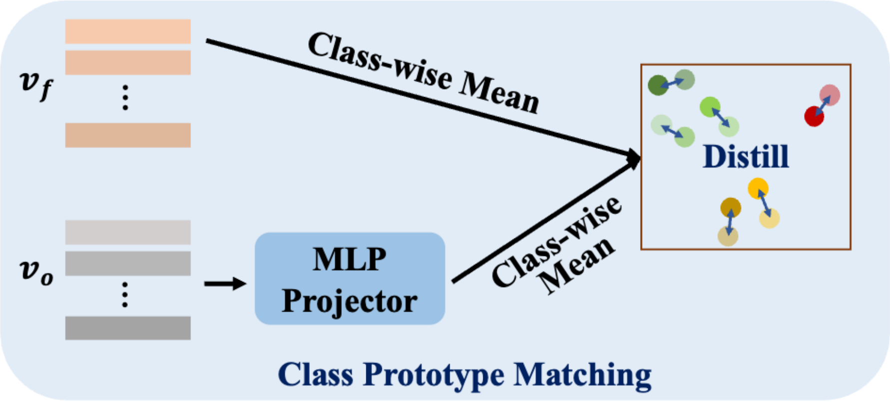


Class Prototype Matching

- Class Prototype
  - Class prototype represents disease-specific features of each class.
  - Computation: During the training per batch, the class prototype vector is the centroid of all the feature vectors belonging to each class.
- Loss function
  - KL divergence is used to encourage OCT student to learn matched class prototypes with fundus teacher.

$$\mathcal{L}_{CPM} = \sum_{c=1}^{C} \mathcal{E}_f^c \log \left(\frac{\mathcal{E}_f^c}{\mathcal{E}_o^c} \right)$$

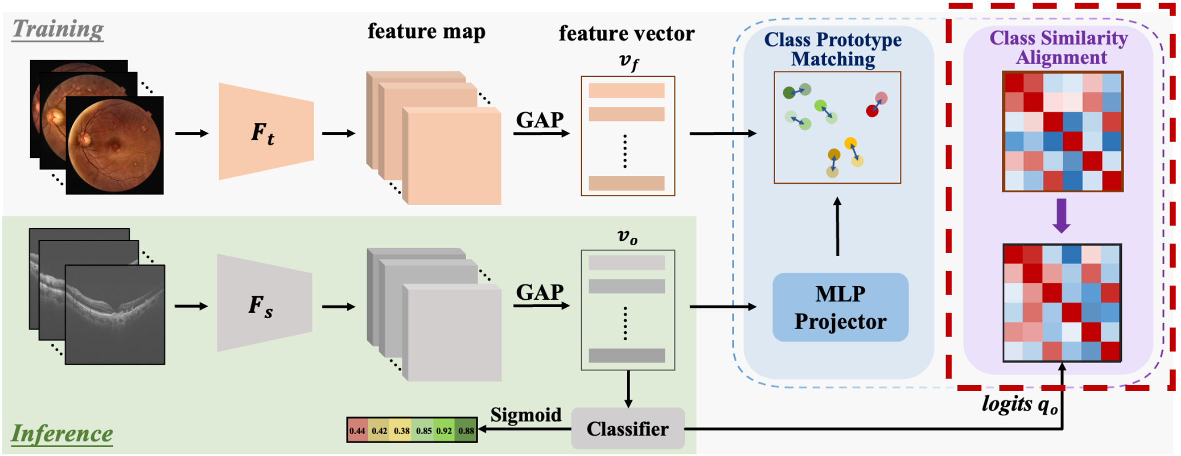


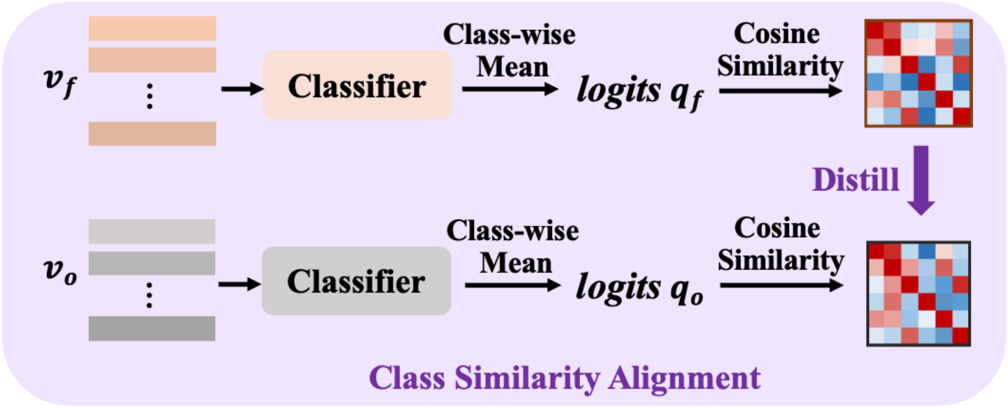


Class Similarity Alignment

- Class Similarity Matrix
  - Class similarity matrix shows correlation between different diseases.
  - Class similarity matrix is based on the cosine similarity of class-wise averaged logits for samples from different classes.
- Loss function
  - KL divergence loss is used to encourage alignment between the two similarity matrices.

$$\mathcal{L}_{CSA} = \sum_{c=1}^{C} \mathcal{Q}_f^c \log \left(\frac{\mathcal{Q}_f^c}{\mathcal{Q}_o^c} \right)$$





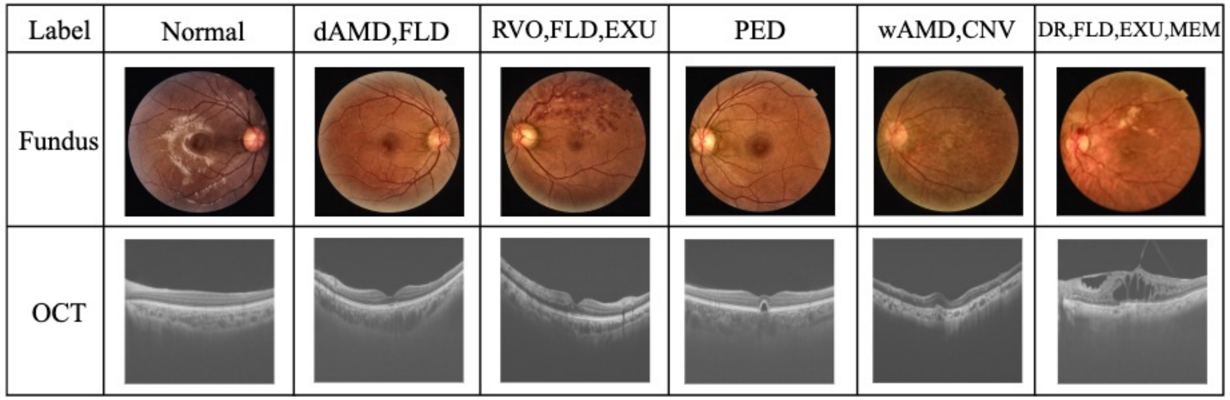
Outline

- Introduction
- Methodology
- Experiment
- Future Challenges
- Conclusion

Dataset

- We collect a new dataset TOPCON-MM with paired fundus and OCT images from 369 eyes of 203 patients.
  - Multiple fundus and OCT images are obtained for each eye.
- Our dataset includes 11 classes.
  - Each image may reveal multiple diseases.
  - Fundus photos and OCT images from each eye have consistent labels.

Category	Normal	dAMD	wAMD	DR	CSC	PED	MEM	FLD	EXU	CNV	RVO	Total
Eyes	153	52	30	72	15	23	38	93	90	14	10	369
Fundus Images	299	178	171	502	95	134	200	638	576	143	34	1520
OCT Images	278	160	145	502	95	133	196	613	573	138	34	1435



Compare with State-of-the-Arts

- Our method outperforms single-modal, multi-modal, and knowledge distillation methods.

	Method	Training	Inference	Paired	MAP	Sensitivity	Specificity	F1 Score	AUC
	Single-Modal Methods								
Single-Modal Learning	ResNet50	Fundus	Fundus	-	50.56±3.05	43.68±6.58	92.24±0.77	54.95±7.09	79.97±1.63
	ResNet50	OCT	OCT	-	66.44±3.81	53.14±6.60	95.28±0.85	64.16±6.24	87.73±1.44
	Multi-Modal Methods								
Multi-Modal Learning	Late Fusion	Both	Both	✓	63.83±1.34	54.45±2.72	94.29±0.74	64.93±3.00	86.92±1.48
	Two-Stream CNN† [25]	Both	Both	✓	58.75±2.71	53.47±3.82	92.97±0.91	61.82±4.02	84.79±2.77
	MSAN† [4]	Both	Both	✓	59.49±3.43	56.44±3.13	93.37±0.59	63.95±3.77	84.51±1.91
Knowledge Distillation	FitNet⋆ [21]	Both	OCT	✓	63.41±3.45	54.44±4.04	94.87±0.58	65.00±4.32	87.17±1.85
	KD⋆ [5]	Both	OCT	✓	63.69±2.04	51.70±3.10	95.75±0.62	63.56±2.32	87.90±1.03
	RKD⋆ [20]	Both	OCT	✓	63.59±3.04	53.42±1.71	94.42±1.81	63.70±0.72	87.36±2.08
	DKD⋆ [27]	Both	OCT	✓	64.40±2.09	53.83±5.23	95.24±0.11	64.00±4.44	87.52±0.58
	SimKD⋆ [1]	Both	OCT	✓	65.10±2.63	53.13±5.49	95.09±0.92	63.19±6.53	87.97±1.32
	Ours	Both	OCT	✗	**69.06**±3.39	**57.15**±5.93	**95.93**±0.57	**69.17**±6.07	**89.06**±0.97

Ablation Study

- CPM improves classification performance in majority classes by distilling disease-specific knowledge.
- CSA benefits minority classes by attending to inter-disease relationships.

Method		MAP			F1 Score	AUC
CPM	CSA	Overall	Majority	Minority		
✗	✗	66.44±3.81	71.12±4.04	58.26±7.42	64.16±6.24	87.73±1.44
✗	✓	67.50±3.00	70.60±4.91	**62.08**±8.72	65.40±5.21	88.09±1.28
✓	✗	67.76±2.34	72.26±3.20	59.90±8.32	65.58±2.58	88.73±1.32
✓	✓	**69.06**±3.39	**73.34**±3.48	61.47±8.17	**69.17**±6.07	**89.06**±0.97

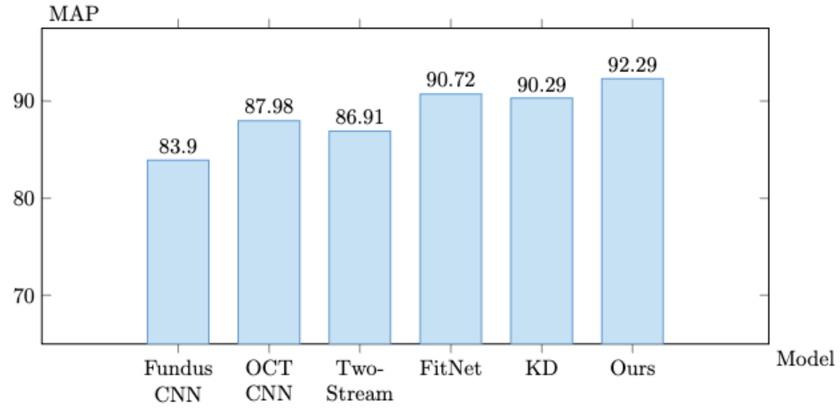
"Majority" and "Minority" respectively refers to the average score of classes that represent more than 10% or less than 10% of the total number of images.

Result on Public Dataset

- MMC-AMD dataset
  - MMC-AMD dataset contains four classes: normal, dry AMD, PCV, and wet AMD.
  - We reproduce single-modal ResNet, Two-Stream CNN and KD methods as baselines.

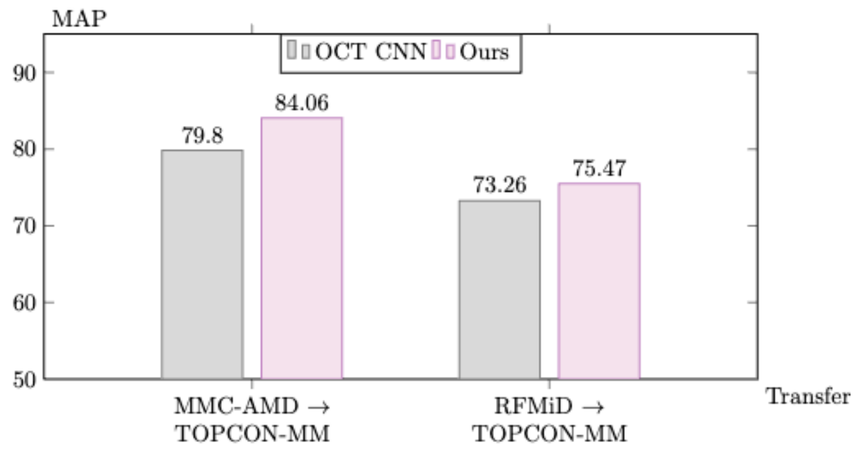
Class	Training set		Validation set		Test set	
	CFP	*OCT*	*CFP*	*OCT*	*CFP*	*OCT*
normal	155 (155)	156 (155)	20 (20)	20 (20)	20 (20)	20 (20)
dryAMD	67 (67)	33 (22)	20 (20)	35 (20)	20 (20)	38 (20)
PCV	259 (259)	289 (156)	20 (20)	44 (20)	20 (20)	47 (20)
wetAMD	453 (452)	531 (325)	20 (20)	38 (20)	20 (20)	38 (20)

[Wang et al, 2022]



Transfer from Other Fundus Datasets

- We separately reproduce our methods with fundus images from two datasets, MMC-AMD and RFMiD. Images from common classes are selected.
- Our method has the flexibility to use any existing fundus dataset to enhance OCT classification.

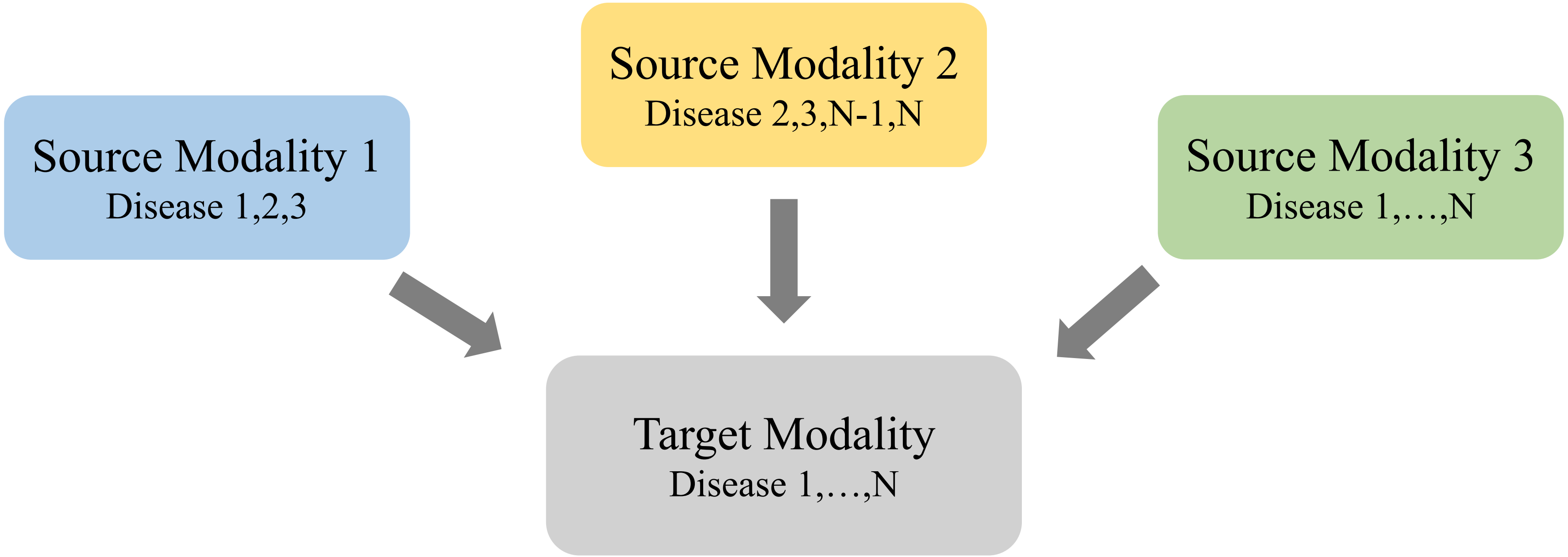


Outline

- Introduction
- Methodology
- Experiment
- Future Challenges
- Conclusion

Future Challenges

- More flexible knowledge transfer from publicly available datasets.
  - Include more modalities (OCTA, etc.)
  - Unaligned label space



Outline

- Introduction
- Methodology
- Experiment
- Future Challenges
- Conclusion

Conclusion

- Our work proposes a novel fundus-enhanced disease-aware distillation module, FDDM, for retinal disease classification.
    - Class prototype matching distills global disease information from the fundus teacher to the OCT student
    - Class similarity alignment ensures the consistency of disease relationships between both modalities.
- Our method reduces the prerequisites for clinical applications.
    - Does not rely on paired data in training stage.
    - Predict with only OCT data in inference stage.
- We make it possible to extract knowledge from any available fundus data.