Class-Conditioned Image Synthesis with Diffusion for Imbalanced Diabetic Retinopathy Grading

MICCAI 2025
28th International Conference on Medical Image Computing and Computer Assisted Intervention
23-27 September 2025
Daejeon Convention Center
Daejeon REPUBLIC OF KOREA

Haochen Zhang [1], Anna Heinke [2], Ines Nagel [2],
Dirk-Uwe Bartsch [2], William Freeman [2], Truong Nguyen [1], Cheolhong An [1]
[1] Electrical and Computer Engineering Department, UC San Diego, USA
[2] Jacobs Retina Center, Shiley Eye Institute, UC San Diego, USA
Email: haz035@ucsd.edu

Diabetic Retinopathy & Fundus Photography

Normal | Mild DR | Moderate DR | Severe DR | Proliferative DR

Mild DR
- Presence of a few microaneurysms
- No significant retinal hemorrhages or vascular abnormalities

Moderate DR
- Increased number of microaneurysms
- Presence of dot-and-blot hemorrhages
- Possible cotton-wool spots and hard exudates

Severe DR
- 4 quadrants of diffuse retinal hemorrhages and microaneurysms
- 2 or more quadrants of venous beading
- 1 or more quadrants of intraretinal microvascular anomalies (IRMA)

Proliferative DR
- Neovascularization of the disc (NVD) greater than 1/4 to 1/3 of the disc area
- Any NVD associated with vitreous or preretinal hemorrhage
- Any neovascularization elsewhere (NVE) associated with vitreous or preretinal hemorrhage

Motivation & Contributions

Motivation

- Public DR datasets (e.g., DDR) suffer from severe class imbalance.
- SOTA classifiers often misclassify underrepresented classes.
- Need for realistic synthetic samples to balance the training set.

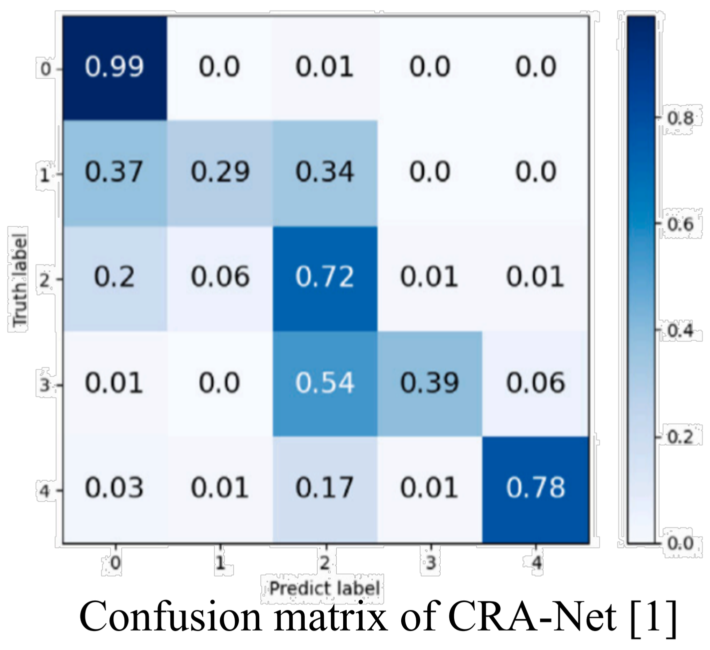
Contribution

- Efficiently fine-tune text-to-image diffusion model on fundus data (Dreambooth for fundus).
- Semantic quality–based evaluation & filtering to ensure effective synthetic training data.
- Explicit class conditioning for high semantic quality sample generation in diffusion training.

Background

Existing Challenges:

- **Imbalanced datasets** → biased training and evaluation
- **High accuracy ≠ reliable diagnosis** (SOTA model [1] fails in mild & severe DR)
- **Attention needed on balanced accuracy** across all classes for clinical usefulness



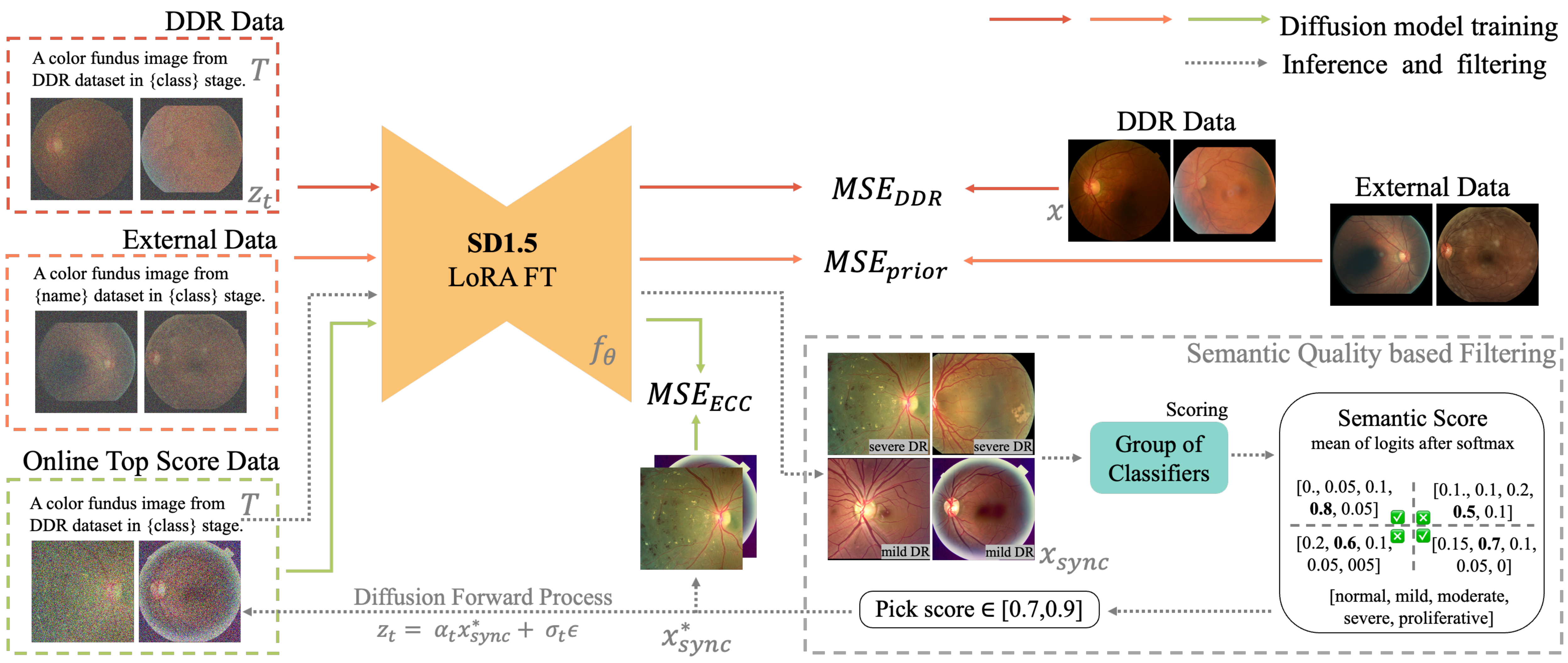
Confusion matrix of CRA-Net [1]

	EyePACS	APTOS	DDR	DDR Test
No DR	25810	1805	6266	1880
Mild DR	2443	370	630	189
Moderate DR	5292	999	4477	1344
Severe DR	873	193	236	71
Proliferative DR	708	295	913	275

Proposed Method & Framework Overview

1. T2I Diffusion Finetuning

- Use MSE_{DDR} as subject-instance loss on fundus dataset.
- Use MSE_{prior} as class-specific prior preservation loss.
- Together form the DreamBooth-style [2] finetuning framework for medical imaging.



2. Semantic Quality–Based Filtering

- Prior work [3, 4] shows more realistic ≠ more useful for classifier training.
- Deploy a committee of pretrained DR classifiers as diagnostic agents.
- Each generated image is scored by the average likelihood across classifiers.
- Low-score samples → degrade training; perfect-score samples → no new information; So pick [0.7, 0.9].

3. Explicit Class Conditioning

Implement in a self-supervised manner

- Stage 1: FT with DreamBooth framework to adapt to fundus domain.
- Stage 2: FT with selected high-semantic-quality samples MSE_{ECC}, together with MSE_{DDR} and MSE_{prior}.

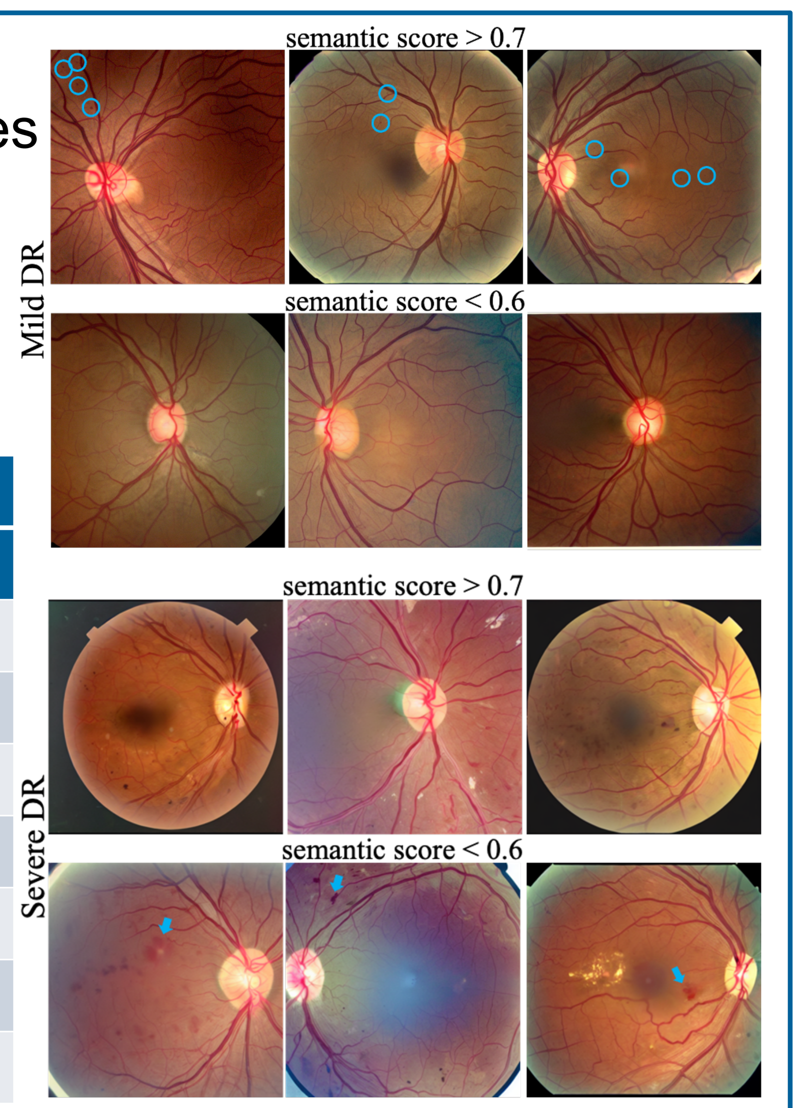
Experiments

Quantitatively ⬇️,

- Oversampling < DM-generated samples → DM helps.
- Basic DM (no filter) = worst performance in DMs.
- Both ECC DMs ≈ basic DM + filter → filtering helps.
- For ECC: no filter > with filter → soft semantic quality control preferred.
- Balanced accuracy ↑, overall accuracy ↓; synthetic data reduces class prior bias.

Backbone	VGG-16			Inception-v3			DenseNet-121		
Metric	B. Acc	Kappa	Acc	B. Acc	Kappa	Acc	B. Acc	Kappa	Acc
LANet [5] *	65.92	**86.41**	**83.83**	67.80	**85.92**	**83.88**	64.24	84.20	82.50
LANet	66.84	85.06	83.67	65.97	84.47	81.70	62.14	**84.26**	**83.08**
+ Oversampling	68.78	82.47	79.06	66.46	84.55	81.28	64.79	83.66	80.69
+ Basic DM w/o Filter	71.37	81.94	76.99	69.01	82.53	78.27	67.19	83.35	78.8
+ Basic DM w/ Filter	73.79	84.24	77.92	71.33	84.18	79.49	69.47	82.32	76.59
+ ECC DM w/o Filter	**74.20**	81.38	75.02	**71.95**	83.76	76.67	**70.22**	83.76	79.76
+ ECC DM w/ Filter	73.82	82.58	76.80	70.38	84.74	77.57	69.20	83.82	79.28

Qualitatively ➡️, high-scoring synthetic images capture key features of specific DR stage.



Conclusion

- Class-conditioned diffusion model mitigates class imbalance in DR grading.
- Semantic filtering ensures useful synthetic samples are retained.
- Proposed framework enables diffusion fine-tuning with limited data and showcases ECC for higher semantic scores.
- Balanced accuracy improved from 66.84% → 74.20%.

Future work

- Expand (rather than balance) datasets with synthetic samples.
- Study interactions between evaluation classifiers and ECC-trained classifiers.

Reference

[1] Zang, F., Ma, H.: CRA-Net: Transformer guided category-relation attention network for DR grading. Computers in Biology and Medicine 170, 107993 (2024)
[2] Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: CVPR. pp. 22500–22510 (2023)
[3] Liu, D., Zhang, H., Xiong, Z.: On the classification-distortion-perception tradeoff. In: NeurIPS 32, 1–10 (2019)
[4] Ye-Bin, M., Hyeon-Woo, N., Choi, W., Kim, N., Kwak, S., Oh, T.H.: Exploiting synthetic data for data imbalance problems: Baselines from a data perspective. arXiv preprint arXiv:2308.00994 (2023)
[5] Hou, J., X., F., Xu, J., Feng, R., Z., Y., Zou, H., Lu, L., Xue, W.: Diabetic retinopathy grading with weakly-supervised lesion priors. In: ICASSP. pp. 1–5 (2023)