A Minimally Supervised Approach for Medical Image Quality Assessment in Domain Shift Settings

Huijuan Yang[1] ✉, Aaron S. Coyner[2], Feri Guretno[1], Ivan Ho Mien[1], Chuan Sheng Foo[1], J. Peter Campbell[2], Susan Ostmo[2], Michael F. Chiang[3] and Pavitra Krishnaswamy[1] ✉
[1]Institute for Infocomm Research, Singapore; [2]Oregon Health & Science University, USA; [3]National Eye Institute, National Institutes of Health, USA

Institute for Infocomm Research
I²R

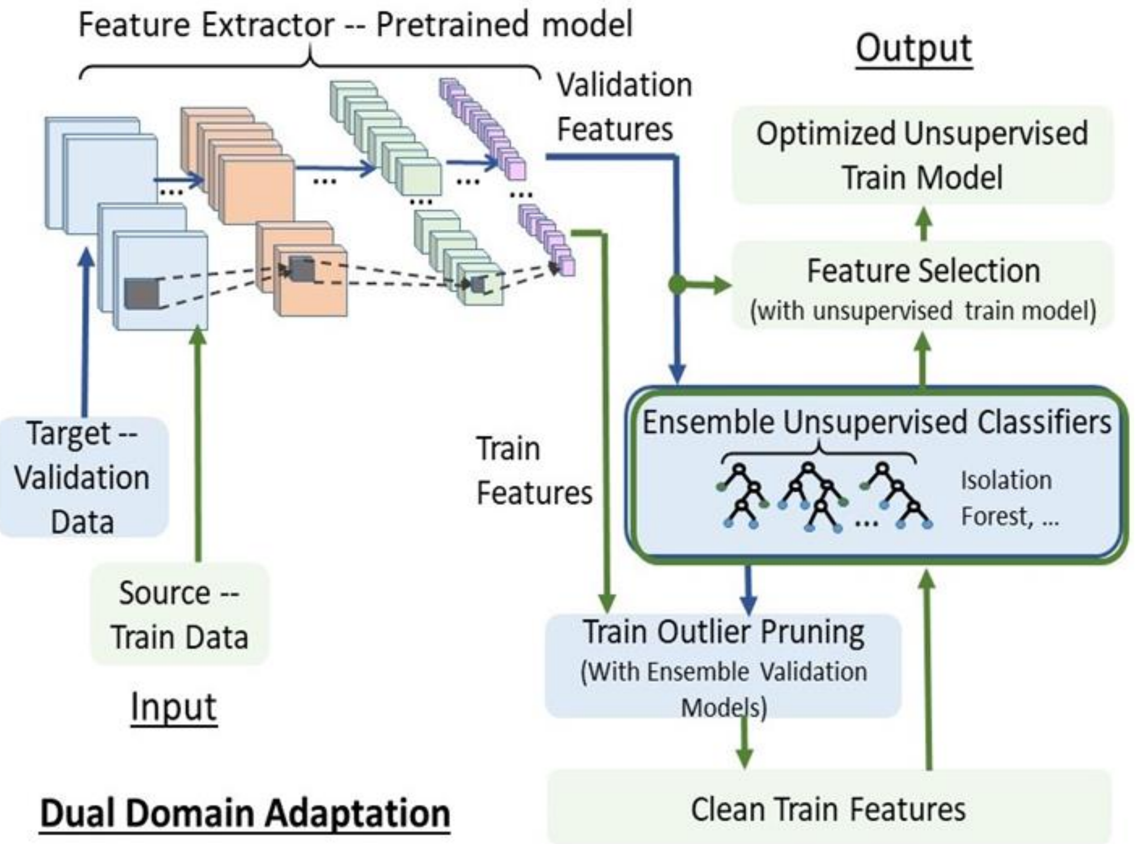
Background and Challenges

- **The Need**
  - Accurate, precise disease diagnosis requires objective image quality assessment (IQA)
  - Automated IQA can identify the need for reacquisition, save time and resources, and enhance screening and diagnosis workflows.
- **Challenges**
  - Data and label scarcity.
    - Difficult to obtain large datasets with varying quality levels to develop IQA approaches
    - Due to dependence of image quality on acquisition setting, need data from different sites for generalizable approaches
    - Deep learning approaches need dedicated quality labels (independent of abnormality itself) from domain experts
  - Class imbalance, noise and artifacts and domain shift. Clinical image datasets exhibit noise/artifacts; class imbalance, and significant domain shifts across acquisition sites.

SOTA-IQA Methods

Methods	Key Techniques	Limitations
Conventional IQA methods for retinal images (Pires Dias et al.'14)	Employ generic parameters and structural parameters	The heavy reliance on identification of anatomical landmarks limits applicability
Deep neural network -based methods (Costa et al. '17, Coyner et al.'18&19, Fu et al.'19)	Extract multi-level features and transfer knowledge to target tasks. Integrates representations of different color-space. Pool the patch classification results	Need large amount of annotated data with varying quality labels
Conventional domain adaptation methods (Lee et al.'19, Morerio et al.'18, Shen et al.'20)	Leverages adversarial dropout, align source and target by geodesic alignment for correlation. Detect optic disc and fovea to assist coarse-to-fine feature encoding	Not perform well for small medical dataset with significant quality variations and distribution shift

Our Proposed Method



- Minimally-supervised image quality assessment (MIQA) approach that learns effectively with small datasets and limited labels in domain shift (DS) scenarios.
- Formulate as anomaly detection task: there is severe class-imbalance (small proportion of images with "unacceptable quality").
- Mitigate DS using a small number of labeled target domain images to identify a compact subset of source domain data with acceptable quality; and use this compact set to train a one-class classifier for IQA.

Experimental Evaluation—Dataset, Performance Comparison

Dataset

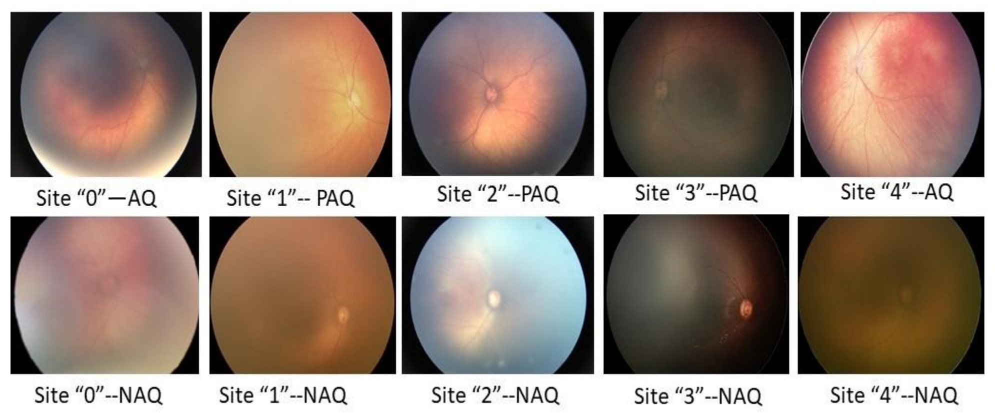
- Real-world multi-center dataset: Retinopathy Of Prematurity image quality (ROP-Quality) [Imaging and Informatics in ROP study (Coyner et al.'18&19)]
- Labeled for diagnostic quality: "Acceptable/Possibly Acceptable/Not Acceptable Quality (AQ/PAQ/NAQ)": consensus rating by 3 annotators.
- 5 sites with 443, 609, 1305, 1475 and 1977 posterior view images→ 20 site pairs

- Public Diabetic Retinopathy image quality dataset (DR-Quality)
- 28,292 DR images, re-annotated with labels of "Good", "Usable", and "Reject".
- Simulate scenarios for small imbalanced datasets, and varying degrees of data and label scarcity.
- Generate 10 splits with 400 to 2500 images each via stratified random sampling.

❖ Designate "Reject" (prevalence 18-21% for DR-Quality) , and "NAQ" (prevalence 1.7-11% for ROP-Quality) classes as the target anomalies for detection.
❖ Randomly sample 3% of the data (i.e., 12-75 images) from target domain while maintaining class proportions for labeling.

Domain Shift in ROP-Quality Dataset

Exemplar ROP images from different acquisition sites
Significant image quality variations across sites



Demographic, socioeconomic, technical and clinical differences (race, ethnicity, birthweight, technical acquisition skills) translate to differences in image features/quality levels across sites

- Empirically quantify degree of domain shift by training a CNN classifier to predict domain class label (classification accuracy related to H-divergence (Ganin et al. JMLR 2016)).
- Across 20 site pairs, the mean classification accuracy was 91.08±4.35%→ significant domain shift amongst acquisition sites.

Performance Comparison

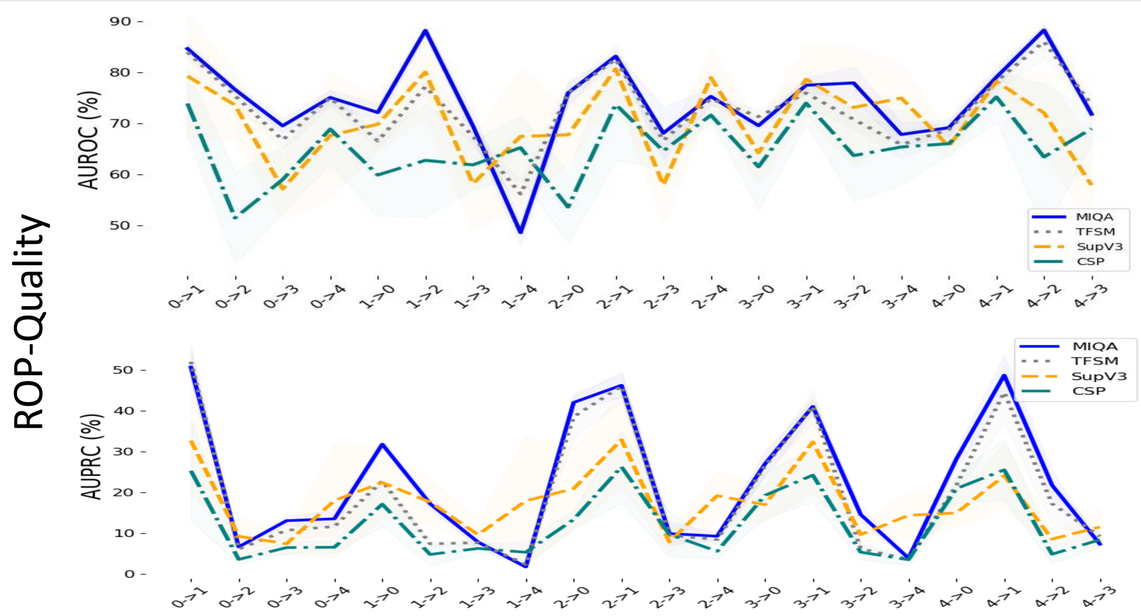
Average Performance Across Multiple Splits and Source-Target Pairs

Dataset	Method	Auroc (%)	Auprc (%)	Gain Over Baseline	
				Auroc (%)	Auprc (%)
DR-Quality	MIQA	90.66±1.14	**75.00±2.70**		
	TFSm	85.27±1.83	62.73± 4.43	+5.39	+12.27
	SupV3	**93.28±1.30**	53.26± 5.08	-2.62	+21.74
	CSP	88.50±2.98	70.81± 5.90	+2.16	+4.19
ROP-Quality	MIQA	**74.37±8.83**	**22.08± 16.17**		
	TFSm	72.93±7.15	19.62± 16.10	+1.44	+2.46
	SupV3	70.13±8.10	17.38± 8.23	+4.24	+4.70
	CSP	65.20±6.58	12.08± 8.45	+9.17	+10.00

TFSm--Transfer Forest with feature selection based on small validation data.
SupV3 and CSP—Supervised baselines based on Inception V3 & Color Space.
MIQA -- Minimally-supervised Image Quality Assessment (proposed)

MIQA far more effective in detecting poor quality images (anomalies)
- DR-Quality: MIQA adapts well to the data scarcity, class imbalance and data variation.
- ROP-Quality: MIQA offers substantial gains across different source-target site pairs.

Experimental Evaluation—Performance Comparison



Selected Layers for Different Site Pairs: ROP-Quality Data

Selected Layer	Site Pairs	Selected Layer	Site Pairs
Mixed_6a	2→0, 2→3, 1→2, 1→0, 3→2, 3→1, 4→2, 4→1	PreAuxLogits	3→0, 4→0
MaxPool_5a_3x3	2→1, 1→3, 0→1, 3→4	Conv2d_2a_3x3	1→4
Mixed_5b	2→4, 0→3,	Conv2d_2b_3x3	0→2
	0→4	Conv2d_3b_1x1	4→3

The performance of MIQA is good even for the real-world multi-center ROP-Quality dataset which exhibits more serious domain shift, site-to-site variation, and class imbalance -> utility in practical settings

MIQA typically adapts the selected layers/ feature representations to nature of target data. Prioritizes lower visual layers when target domain has more data; and higher semantic layers when data in target domain more scarce.

Summary

- We presented a Minimally-supervised Image Quality Assessment (MIQA) method for medical images.
- MIQA uses an anomaly detection framework to collectively address data and label scarcity, class imbalance, and domain shift across acquisition sites.
- MIQA employs a small target validation dataset to improve representation of features pertaining to images of acceptable quality, and then leverages a one-class classifier to detect images of poor quality.
- In experiments on multi-center medical image quality datasets, we demonstrate large performance gains over existing supervised and semi-supervised baselines.
- Our work has implications as a tool for improved image quality audit in many clinical settings and AI deployment applications.

Contact and Acknowledgment

✉ hjyang@i2r.a-star.edu.sg, pavitrak@i2r.a-star.edu.sg

Research efforts were supported by funding and infrastructure for deep learning and medical imaging research from the Institute for Infocomm Research, Science and Engineering Research Council, A*STAR, Singapore. For experimental evaluations on real-world data, we gratefully acknowledge grants R01EY19474, R01 EY031331, and P30 EY10572 from the National Institutes of Health (Bethesda, MD), and by unrestricted departmental funding and a Career Development Award (JPC) from Research to Prevent Blindness (New York, NY). We also acknowledge insightful discussions with Vijay Chandrasekhar from the Institute for Infocomm Research and Jayashree Kalpathy-Cramer from the Massachusetts General Hospital, Boston USA.