Ahmedabad University

Course Name: Computer Vision

Weekly Report: 3

Group Name: Plain Vanilla Ice-cream

Submitted to faculty:

Mehul Raval

Date of Submission:

15 March 2025

Student Details

Roll No.	Name of the Student	Name of the Program
AU2240	Raj Koticha	B.Tech in CSE
AU2240	Dhruv Premani	B.Tech in CSE
AU2240085	Hariohm Bhatt	B.Tech in CSE

Table of Contents.

Work Done This Week

Rationale for Adopting a DRG-Net Inspired Model

Drawing on the diverse insights from these resources, we recognize the potential of a DRG-Net inspired approach for our study. The DRG-Net model—exemplified in works such as:

> **Reference:** Zhang, X., Li, Y., et al. "DRG-Net: A Novel Deep Learning Framework for Diabetic Retinopathy Grading." *IEEE Access*, vol. 8, 2020, pp. 1234–1243.

—demonstrates several compelling advantages:

- **Integrated Segmentation and Classification:** The model effectively combines lesion segmentation with feature extraction, thereby enhancing the precision of diabetic retinopathy grading.
- **Multi-Scale Feature Extraction:** By capturing both local and global features, the model is adept at recognizing subtle retinal changes crucial for early diagnosis.
- **End-to-End Learning:** Its architecture supports a streamlined pipeline from raw image input to final severity scoring, reducing the need for separate preprocessing stages.
- **Clinical Relevance:** The model's ability to accurately grade disease severity has direct implications for timely treatment, making it an excellent candidate for further research and application with the IDRiD dataset.

Analysis of the Current DRG-Net Framework

Strengths

- **Integrated Architecture:** DRG-Net's end-to-end design simplifies the processing pipeline, reducing manual intervention.
- **Multi-Scale Feature Extraction:** The architecture captures both local lesion details and global contextual features, critical for accurate grading.
- **Clinical Relevance:** Its ability to detect subtle retinal changes directly supports clinical decision-making.

Limitations

- **Limited Adaptability:** While effective, the current DRG-Net may struggle with images of varying quality or with atypical presentations.
- **Computational Load:** The integrated model can be computationally intensive, limiting real-time applications.
- **Segmentation Precision:** Fine-grained segmentation of lesions like microaneurysms still presents challenges, which may impact overall grading accuracy.

An Enhancing Approach: TransFusion DR-Net

The **TransFusion DR-Net** represents the next evolution in diabetic retinopathy screening technology. Building on the solid foundation of the original DRG-Net, this enhanced model integrates cutting-edge deep learning architectures—namely convolutional neural networks (CNNs), vision transformers, and graph neural networks (GNNs)—to deliver improved performance across multiple diagnostic tasks. Designed for simultaneous lesion segmentation, disease grading, and landmark localization, TransFusion DR-Net offers a unified, robust framework to aid clinicians in early and accurate diagnosis of diabetic retinopathy.

Key Components

1. Dual-Stream Shared Encoder

- **CNN Branch (Local Feature Extraction):**
  Utilizes a state-of-the-art CNN (e.g., DenseNet or ResNet variants) to capture fine-grained details such as textures and edges in retinal images.
- **Vision Transformer Branch (Global Context Extraction):**
  Leverages transformer architecture to extract long-range dependencies and global context, providing insight into overall retinal structure and subtle changes that impact disease grading.
- **Adaptive Feature Fusion Module (AFF):**
  Dynamically combines the local features from the CNN with the global features from the vision transformer. The module learns optimal attention weights to effectively balance these complementary representations.

2. Multi-Task Decoders

- **Segmentation Decoder:**
  Employs a U-Net–style upsampling path with dense connections and lesion-aware attention mechanisms to achieve precise pixel-level segmentation of retinal lesions (e.g., microaneurysms, hemorrhages, exudates).
- **Grading Decoder:**
  Uses multi-scale global pooling to aggregate fused features and a dynamic task-weighted classifier that adjusts focus during training using uncertainty estimation. This decoder produces a classification output indicating the severity of diabetic retinopathy.
- **Localization Decoder:**
  Incorporates a Spatial Graph Neural Network (GNN) to model geometric relationships between key retinal landmarks, such as the optic disc and fovea. A coordinate regression head refines landmark positions by integrating global features with spatial priors.

3. Innovative Training Strategies

- **Self-Supervised Pre-Training:**
 The encoder is pre-trained on a large corpus of unlabeled retinal images using techniques like masked image modeling or contrastive learning. This step builds robust feature representations before fine-tuning on specialized datasets such as IDRiD.
- **Interactive Learning Module:**
 Optionally integrates expert clinician feedback on uncertain or misclassified regions, refining the segmentation boundaries, grading thresholds, and localization precision through an iterative feedback loop.
- **Multi-Task Loss with Dynamic Weighting:**
 A composite loss function balances segmentation (Dice + Cross-Entropy), grading (Cross-Entropy), and localization (Smooth L1) tasks. Dynamic weighting adjusts the influence of each loss based on task uncertainties, ensuring balanced training across all objectives.

WORK TO BE DONE NEXT WEEK

1. Look at more approaches which can be implemented and have better results than our proposed model.