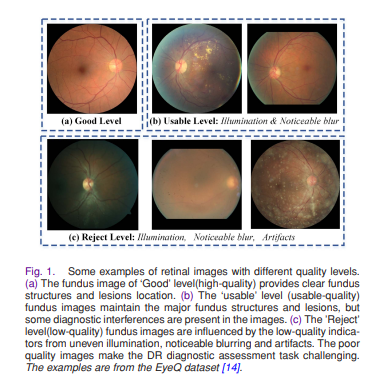
**ABSTRACT**

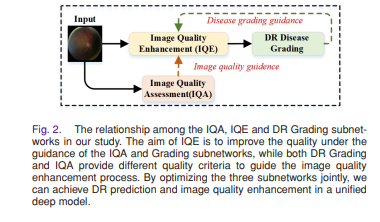
Diabetic retinopathy is a serious eye condition affecting individuals with diabetes and can lead to vision impairment or even blindness if not detected and treated in its early stages. In this project, we present a novel approach to automate the grading of diabetic retinopathy using deep learning techniques, implemented in Matlab. Our developed system leverages a Convolutional Neural Network (CNN) model architecture to achieve a remarkable accuracy rate of 92%, offering a reliable and efficient solution for diabetic retinopathy diagnosis. The proposed system consists of a multi-stage process. In the initial stage, retinal images are inputted into the system. The second stage encompasses preprocessing, which plays a pivotal role in enhancing the quality of the input images. It involves the removal of noise using a median filter and the improvement of image contrast through the Contrast Limited Adaptive Histogram Equalization (CLAHE) process. The heart of our system lies in the third module, where retinal images are subjected to classification based on the severity of diabetic retinopathy. Our CNN model, finely tuned with hyperparameters including epochs, learning rate, dropout rate, and optimizer (Stochastic Gradient Descent with Momentum - SGDM), is trained to make accurate predictions. It classifies retinal images into distinct grades, including 'no apparent retinopathy' (grade 0), 'mild NPDR' (grade 1), 'moderate NPDR' (grade 2), 'severe NPDR' (grade 3), and 'Proliferative Diabetic Retinopathy' (grade 4). This grading system aids in early diagnosis and timely intervention. To assess the performance of our proposed model, we employ a comprehensive set of evaluation metrics, including accuracy, error rate, precision, recall, specificity, F1-score, and Matthews Correlation Coefficient (MCC). These metrics provide a holistic evaluation of the system's performance, ensuring its reliability and effectiveness in diabetic retinopathy grading. In conclusion, our project represents a significant advancement in the automated diagnosis of diabetic retinopathy, demonstrating the power of deep learning and CNNs in medical image analysis. With an impressive accuracy of 92%, our system holds great promise for improving the early detection and management of diabetic retinopathy, thereby enhancing the quality of life for individuals living with diabetes.

**INTRODUCTION**

DIABETIC retinopathy (DR) is one of the most serious complications of diabetes and is currently the leading cause of blindness in adults [1], and has been identified by the World Health Organization as the second most serious eye disease after cataract. Owing to the safety and cost-effectiveness of acquiring fundus images, they are widely used for early screening and diagnosis of DR [2], [3]. However, due to the limitations of the acquisition equipment and the operation procedure, the fundus images often present significant differences with respect to the image quality as shown in Fig. 1. Automatic identification of lesions, such as microaneurysms (MAs) and hard exudates (EXs) are crucial to the diagnostic assessment of DR. The lower quality leads to the failure of the identification of the suspicious lesions, which decreases the diagnosis performance. Therefore, it is desirable to enhance the image quality for accurately capturing the lesions related to the severity grading. Disease grading and image quality enhancement are two main fundamental tasks in this area. The image quality enhancement is required to be guided by the image quality assessment (IQA), the aim of which is to measure and control the quality of images.

However, IQA is a subjective task depending on the experience of the ophthalmologists. The current solution turns it into learning, data-driven approaches based on neural networks. Through the thorough analysis of disease grading, image quality enhancement and quality assessment in Section IV, we believe that a joint framework incorporating the image quality assessment, image quality enhancement, and disease grading is feasible and significant, but to the best of our knowledge, no such work has been studied in this field. The challenges mainly lie in: (I) how to appropriately evaluate the quality of fundus images; (II) how to effectively enhance the low-quality fundus images, and (III) how to develop an end-to-end collaborative learning framework by integrating image quality enhancement subnetwork, image quality assessment subnetwork and DR subnetwork. To solve these issues, we propose an image Quality Assessment guided Collaborative Learning framework for both image quality Enhancing and DR grading, called CLEAQ-DR. The framework takes into account the image quality assessment and image quality enhancement during the DR grading. The underlying assumption is that under the guidance and help of the quality assessment and enhancement, the lesion’s identification capability and DR grading performance can be improved. To better explore the potential relationships among the components of the quality assessment, the quality enhancement and the grading in the fundus images, we propose a collaborative learning framework to explore the potential correlation among these tasks and jointly train these subnetworks in a unified deep model for improving the individual performance. To this end, our collaborative learning framework incorporates three subnetworks: a DR disease grading subnetwork for predicting the DR level, a two-branch image quality enhancement (IQE) subnetwork for improving the image quality while preserving the fundus structure, and a two-branch image quality assessment (IQA) subnetwork for capturing the inherent lowquality indicators and predicting the quality level. Specifically, the IQE.





subnetwork consists of two encoder-decoder modules, where an image quality enhancement (U-IQE) module aims to learn the mapping relationship of low-quality images to high-quality images, and a retina vessel structure segmentation (Seg-IQE) module that aims to model the vessel structure to guarantee the preservation of the main fundus structure during the enhancement procedure. Moreover, the IQA subnetwork involves a classification (C-IQA) module for producing a reliable quality level, and an encoder-decoder (LQI-IQA) module for capturing the critical low-quality indicators by reconstructing the input images into the low-quality images. The image quality assessment, image quality enhancement and disease grading tasks are optimized in an end-to-end manner. There are three notable characteristics for the CLEAQ-DR on the fundus retinal images.

1) Modeling task relationship: There exist inherent relationships among image quality assessment, image quality enhancement and DR grading tasks. Fig. 2 illustrates the inherent relationships among the three subnetworks in our framework. Appropriately modeling the task relationship allows to improve the performance of each task.

2) Exploiting the image quality from different aspects: To comprehensively guide IQE to improve the image quality, the DR and IQA tasks focus on the image quality from the lesion level and the global image level.

3) Collaborative Learning: To better reinforce each task, it is necessary to jointly train the DR grading, IQA and IQE tasks within a unified framework. With such an end-to-end trainable framework, our study establishes the association among the tasks of image quality assessment, enhancement and DR grading by collaborating the three subnetworks for better recovering the image quality and facilitating the precise localization of lesions. In summary, our contributions can be summarized as follows. A major limitation of most current automatic DR grading models is that they ignore the effect of the image quality on the grading performance. To the best of our knowledge, our work is the first attempt to simultaneously perform multiple tasks image quality assessment, image quality enhancement and disease diagnosis through an end-to-end collaborative learning framework. Considering that image quality assessment is essential for DR grading, our study establishes the association among the quality assessment, the quality enhancement and DR disease grading.

We propose a two-branch encoder-decoder image enhancement subnetwork for improving low-quality images while preserving major retinal structures for avoiding the distortion occurrence, which helps to improve the DR diagnosis performance. Moreover, we propose a two-branch image quality assessment subnetwork for assessing the quality and guiding the enhancement process. The module can learn the inherent low-quality indicators for enhancing the assessment performance. Both subnetworks can be easily extended to other tasks related to the low-quality medical images.

Experiment results on two benchmark datasets (Messidor and EyeQ) demonstrate that our approach leads to a significant performance boost over existing networks for DR grading and quality enhancement, notably on the EyeQ dataset that contains a large number of low-quality images. Moreover, we thoroughly analyze the potential correlation among the tasks of image quality assessment, image quality enhancement and DR grading through a series of experiments, demonstrating that these three tasks can benefit from each other. The proposed joint learning framework CLEAQ-DR can be broadly applied to other tasks of medical images with low-quality in general. To make the framework more general, we also further evaluate the enhanced results of low-quality images in more downstream tasks, such as vessel segmentation.

**WHAT IS MEDICAL IMAGE PROCESSING?**

[Medical image processing](https://www.synopsys.com/simpleware/clinical-applications.html) encompasses the use and exploration of 3D image datasets of the human body, obtained most commonly from a Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) scanner to diagnose pathologies or guide medical interventions such as [surgical planning,](https://www.synopsys.com/simpleware/resources/case-studies/patient-specific-thr.html) or for research purposes. Medical image processing is carried out by radiologists, engineers, and clinicians to better understand the anatomy of either individual patients or population groups.

What are the benefits of medical image processing?

The main benefit of medical image processing is that it allows for in-depth, but non-invasive exploration of internal anatomy. [3D models](https://www.synopsys.com/simpleware/human-body-models.html) of the anatomies of interest can be created and studied to improve treatment outcomes for the patient, develop improved medical devices and drug delivery systems, or achieve more informed diagnoses. It has become one of the key tools leveraged for medical advancement in recent years.

The ever-improving quality of imaging coupled with advanced software tools facilitates accurate digital reproduction of anatomical structures at various scales, as well as with largely varying properties including bone and soft tissues. Measurement, statistical analysis, and creation of simulation models which incorporate real anatomical geometries provide the opportunity for more complete understanding, for example of interactions between [patient anatomy and medical devices](https://www.synopsys.com/simpleware/news-and-events/webinars-AI-medical-devices.html).

How does medical image processing work?

The process of medical image processing begins by acquiring raw data from CT or MRI images and reconstructing them into a format suitable for use in relevant software. A 3D bitmap of greyscale intensities containing a voxel (3D pixels) grid creates the typical input for image processing. CT scan greyscale intensity depends on X-ray absorption, while in MRI it is determined by the strength of signals from proton particles during relaxation and after application of very strong magnetic fields.

For medical users, the reconstructed image volume is typically processed to segment out and edit different regions of anatomical interest, such as tissue and bone. In [Synopsys Simpleware software](https://www.synopsys.com/simpleware.html), for example, users can carry out different image processing operations at the 2D and 3D level, including:

* Reducing and removing unwanted noise or artifacts with image filters
* Cropping and resampling input data to make it easier and faster to process images
* Using segmentation tools to identify different anatomical regions, including automated techniques using AI-based machine learning algorithms
* Applying measurement and statistics tools to quantify different parts of the image data, for example, centrelines
* Importing CAD models, such as implants or medical devices, to study how they interact with individual anatomies
* Exporting processed models for physics-based simulation, further design work, or for 3D printing physical replicas of the anatomy in question

## Where and when does medical image processing fit in the product portfolio?

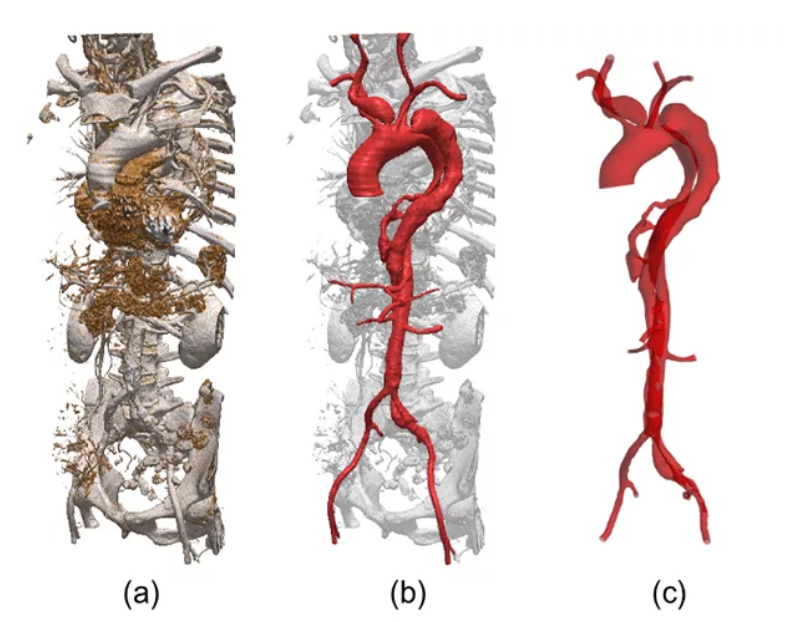
Simpleware software has extensive medical applications, from [general research](https://www.synopsys.com/simpleware/life-sciences.html) to [clinical workflows](https://www.synopsys.com/simpleware/clinical-applications.html) that come under [FDA 510(k) and CE-marking certifications](https://www.synopsys.com/simpleware/resources/regulatory.html). In general, the software provides multiple ways to work with MRI, CT, and other forms of medical image data, including the ability to easily create models that include [CAD-designed implants and devices](https://www.synopsys.com/simpleware/life-sciences/orthopedics.html). Users such as device engineers apply the software to problems like planning surgical procedures, and assessing the performance of different implant designs through tools in Simpleware ScanIP, as well as export of models for simulation and design.

Going beyond medical image processing

Several additional modules are available with [Simpleware ScanIP](https://www.synopsys.com/simpleware/software/scanip.html) to do more with medical image data after initial processing. In addition, options are available for customizing steps and automating repetitive or time-consuming tasks. For example, medical users can:

* Export STL files from processed medical images for [3D printing](https://www.synopsys.com/simpleware/life-sciences/medical-models.html)
* Combine [CAD-designed](https://www.synopsys.com/simpleware/software/cad-module.html) implants with anatomical image data for sizing and positioning
* Generate [volume meshes](https://www.synopsys.com/simpleware/software/fe-module.html) for Finite Element and Computational Fluid Dynamics simulation of physics, such as impact or stress and strain
* Continue design work by converting processed image data into CAD-friendly [NURBS](https://www.synopsys.com/simpleware/software/nurbs-module.html), as well as communicating with [leading CAD packages](https://www.synopsys.com/simpleware/software/design-link-module.html) when developing products
* Use a range of [AI-enabled, automated off-the-shelf software tools and customized solutions](https://www.synopsys.com/simpleware/software/auto-segmenter-modules.html) for speeding up common medical image processing workflows

Putting Medical Image Processing into Practice



*Segmentation of aortic dissection: (a) rendering of the CT data; (b) segmented mask after smoothing; (c) 3D model used in the simulation*

A good recent example of how medical image processing involved patient-specific hemodynamic simulations of [complex aortic dissections](https://www.synopsys.com/simpleware/resources/case-studies/aortic-dissections.html), part of work carried out at [University College London](https://www.ucl.ac.uk/multiscale-cardiovascular-engineering/) into better understanding life-threatening vascular conditions. Researchers used Simpleware software to process CT scans and build models suitable for CFD analysis, with the following steps taken:

**1. CT scans**are obtained from patient-specific cases of aortic dissections

**2. Data** is imported to Simpleware ScanIP to reconstruct patient geometry, including the processing of noise, and segmentation of regions of interest such as the dissected aorta and branches

**3. Scripting** is used to automatically carry out smoothing algorithms to remove pixelation artifacts

**4. Surface models** are generated from the dissected aorta and imported to ANSYS® software to set-up CFD simulations, including intraluminal pressure and wall shear-stress-based indices,

**5. Simulation results**create hemodynamic insights that can be used to help future clinical understanding

**EXISTING SYSTEM**

* The existing system for grading diabetic retinopathy relied solely on image processing techniques to assess the severity of the condition. This traditional approach aimed to diagnose diabetic retinopathy by analyzing retinal images without the use of deep learning or machine learning models. While this system did not incorporate the advantages of modern neural networks, it played a vital role in early detection and provided a foundation for the development of more advanced methods.
* In the existing system, the workflow began with the input of retinal images, just like the proposed system. However, instead of leveraging deep learning, it primarily relied on a series of image processing techniques to extract relevant features and make diagnostic assessments.
* The first stage of the earlier system involved preprocessing, which was vital for improving the quality of retinal images. Techniques such as image denoising, contrast enhancement, and resizing were applied to ensure that the input images were suitable for subsequent analysis. Image denoising aimed to remove unwanted noise from the retinal images, while contrast enhancement techniques enhanced the visibility of important structures within the image. Resizing was often performed to standardize the images' dimensions for consistency in analysis.
* Following preprocessing, the system utilized various image processing algorithms to extract features indicative of diabetic retinopathy. These algorithms were designed to identify and quantify specific characteristics such as microaneurysms, hemorrhages, exudates, and vascular changes within the retinal images. These features were critical for grading the severity of diabetic retinopathy.
* The grading process in the existing system relied on predefined rules and thresholds, which were established based on medical expertise and clinical experience. These rules helped categorize retinal images into different grades, typically ranging from 'no apparent retinopathy' to 'severe proliferative diabetic retinopathy.' The grading system aimed to assess the level of damage to the retinal blood vessels, thus guiding treatment decisions.
* While the existing system lacked the sophistication and accuracy of deep learning models, it played a crucial role in the early diagnosis of diabetic retinopathy and served as a foundation for subsequent developments. The reliance on image processing techniques was a significant step forward in the automation of diabetic retinopathy grading, albeit with limitations in accuracy and adaptability compared to modern deep learning approaches.
* In summary, the existing system for grading diabetic retinopathy through image processing techniques served as a valuable initial approach to automate the diagnosis of this condition. While it lacked the advanced capabilities of deep learning, it laid the groundwork for the development of more accurate and efficient grading systems, ultimately contributing to the improvement of diabetic retinopathy diagnosis and patient care.

**DISADVANTAGES OF EXISTING SYSTEM**

* Limited Diagnostic Accuracy: One of the primary drawbacks of the existing system is its limited diagnostic accuracy. Relying solely on image processing techniques, the system may struggle to accurately detect and classify subtle or complex features of diabetic retinopathy. This limitation can lead to misdiagnosis or the failure to detect early-stage retinopathy.
* Sensitivity to Image Quality: The performance of the system is highly sensitive to the quality of input retinal images. Variations in factors such as image resolution, lighting conditions, and the presence of artifacts can significantly impact the system's ability to make accurate assessments. Inconsistencies in image quality can lead to unreliable results.
* Lack of Adaptability: The existing system lacks adaptability and robustness. It often relies on predefined rules and thresholds, which may not be suitable for handling variations in retinal pathology or accommodating different imaging setups. This lack of adaptability limits the system's ability to handle diverse clinical scenarios effectively.
* Difficulty in Handling Complex Cases: Diabetic retinopathy can manifest in various forms, including subtle microaneurysms to severe hemorrhages and proliferative changes. The existing system may struggle to handle complex cases that involve multiple pathologies or a combination of symptoms, as it may not have the capacity to discern intricate patterns.
* Dependency on Expert Knowledge: The system heavily depends on expert knowledge for designing image processing algorithms and setting diagnostic thresholds. This dependency can make it challenging to update and adapt the system as medical knowledge evolves or as new insights into diabetic retinopathy emerge.
* Inability to Learn and Improve: Unlike modern machine learning and deep learning approaches, the existing system lacks the ability to learn from data and improve its performance over time. It does not have the capacity to adapt to new trends or continuously refine its diagnostic capabilities.
* Limited Scalability: Scaling the existing system to handle large datasets or to integrate with electronic health records systems can be challenging. It may not have the scalability required to support the increasing demand for diabetic retinopathy screening in a clinical setting.
* Inefficient Workflow: The reliance on manual image processing and grading steps in the existing system can be time-consuming and labor-intensive. This inefficiency can lead to delays in diagnosis and treatment, which may impact patient outcomes, especially in cases requiring urgent intervention.
* Subjectivity in Grading: The reliance on predefined rules and thresholds introduces an element of subjectivity into the grading process. Different experts or practitioners may have varying interpretations, leading to inconsistencies in diagnosis and treatment recommendations.
* Lack of Integration: The existing system may not seamlessly integrate with other healthcare information systems or electronic health records, hindering the seamless sharing of patient data and collaboration among healthcare professionals.
* In conclusion, while the existing system for diabetic retinopathy grading through image processing paved the way for automated diagnosis, it has several notable disadvantages, including limited accuracy, sensitivity to image quality, and an inability to adapt to evolving medical knowledge. These shortcomings highlight the need for more advanced and adaptable approaches, such as deep learning models, to improve the accuracy and efficiency of diabetic retinopathy diagnosis.

**PROPOSED SYSTEM**

* The proposed system represents a significant advancement in the field of diabetic retinopathy diagnosis by incorporating deep learning techniques, specifically a Convolutional Neural Network (CNN) model architecture. Unlike the existing system, which relies solely on image processing, this novel approach integrates machine learning to enhance the accuracy and efficiency of diabetic retinopathy grading.
* In the proposed system first the retinal images were obtained from a public database. Image pre-processing is a crucial stage in the image analysis process. Pre-processing is the process of removing unnecessary information from images and improving their quality. We propose a median filter for noise removal and contrast limited adaptive histogram equalization for image enhancement to accomplish this.
* Once the preprocessing is complete, the training phase can begin. Throughout the training phase, images are trained as feature maps using the CNN model. The cornerstone of the proposed system is the integration of a CNN model architecture. This deep learning approach allows the system to learn and extract intricate features from retinal images, improving its ability to discern subtle and complex pathologies associated with diabetic retinopathy. The CNN model acts as a powerful feature extractor and classifier, significantly enhancing the diagnostic capabilities of the system.
* By leveraging deep learning, the proposed system offers the potential for significantly improved diagnostic accuracy. The CNN model can identify patterns and features in retinal images that may not be easily discernible through traditional image processing techniques. This enhancement in accuracy is crucial for early detection and appropriate treatment planning.
* Once the preprocessing is complete, the training phase can begin. Throughout the training phase, images are trained as feature maps using the CNN model. The proposed system excels in grading precision, as it is trained on a diverse dataset with labeled retinal images. The CNN model can accurately categorize images into different diabetic retinopathy grades, including 'no apparent retinopathy,' 'mild NPDR,' 'moderate NPDR,' 'severe NPDR,' and 'Proliferative Diabetic Retinopathy.' This precise grading is essential for tailored treatment plans.
* In summary, the proposed system for diabetic retinopathy grading leverages deep learning, specifically a CNN model, to revolutionize the accuracy and efficiency of diagnosis. This approach offers improved diagnostic precision, adaptability, reduced subjectivity, and the potential for ongoing enhancement, making it a promising solution for enhancing diabetic retinopathy diagnosis and patient care.

**ADVANTAGES OF PROPOSED SYSTEM**

* Enhanced Diagnostic Accuracy: The incorporation of a Convolutional Neural Network (CNN) model enables the proposed system to achieve significantly higher diagnostic accuracy compared to traditional image processing techniques. It can detect subtle and complex features associated with diabetic retinopathy, reducing the risk of misdiagnosis.
* Early Detection: The high accuracy of the proposed system allows for the early detection of diabetic retinopathy, even in its nascent stages. Early diagnosis facilitates timely intervention and treatment, potentially preventing vision loss and improving patient outcomes.
* Automatic Feature Extraction: Unlike conventional methods that require manual feature engineering, the proposed system automatically learns and extracts relevant features from retinal images. This eliminates the need for expert knowledge in feature selection and ensures the system's adaptability to diverse cases.
* Objective Grading: The proposed system provides objective and consistent grading results. It eliminates the subjectivity associated with human interpretation and predefined rules, ensuring that diagnoses are based on learned patterns and features.
* Multi-Class Classification: The CNN model can classify retinal images into multiple diabetic retinopathy grades, including 'no apparent retinopathy,' 'mild NPDR,' 'moderate NPDR,' 'severe NPDR,' and 'Proliferative Diabetic Retinopathy.' This comprehensive grading system offers a nuanced assessment of disease severity.
* Adaptability: Deep learning models are adaptable and can evolve with changes in medical knowledge and diagnostic criteria. The proposed system can continually improve its performance as more data becomes available and as the understanding of diabetic retinopathy advances.
* Efficiency: Deep learning models, once trained, can efficiently process large volumes of data. The proposed system can handle extensive datasets and high-throughput clinical environments, expediting the diagnostic process and reducing delays in treatment.Reduced Workload: By automating the diagnostic process, the proposed system reduces the workload on healthcare professionals, allowing them to focus on patient care and treatment decisions rather than manual image analysis.
* Scalability: The system can easily scale to accommodate the increasing demand for diabetic retinopathy screening. It can be integrated into telemedicine platforms, enabling remote screening and diagnosis in underserved areas.
* Continuous Improvement: The proposed system can continuously improve its diagnostic capabilities through iterative model training and validation. This ongoing enhancement ensures that it remains at the forefront of diabetic retinopathy diagnosis.
* Research and Education: The availability of a robust deep learning model in the proposed system can be invaluable for research and education in the field of diabetic retinopathy. It can serve as a valuable tool for studying disease progression and improving medical education.
* In conclusion, the proposed system offers numerous advantages, including improved diagnostic accuracy, early detection, automatic feature extraction, objectivity, adaptability, efficiency, reduced workload, scalability, and the potential for continuous improvement. These advantages make it a promising solution for enhancing diabetic retinopathy diagnosis, ultimately benefiting patients and healthcare providers alike.

**LITERATURE REVIEW**

**IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021** To provide up-to-date estimates of undiagnosed diabetes mellitus (UDM) prevalence - both globally, and by region/country, for the year 2021.Data sources reporting diabetes prevalence were identified through a systematic search in the peer-reviewed and grey literature. The prevalence of undiagnosed diabetes was estimated from the data from each country where data was available. For countries without in-country data, the prevalence of undiagnosed diabetes was approximated by extrapolating the average of the estimates from countries with data sources within the same International Diabetes Federation (IDF) region and World Bank income grouping. We then applied these stratified prevalence estimates of UDM from each country to the number of adults in each strata and summed the counts to generate the number of adults with UDM (aged 20–79 years) for 215 countries and territories.In 2021, almost one in two adults (20–79 years old) with diabetes were unaware of their diabetes status (44.7%; 239.7 million). The highest proportions of undiagnosed diabetes (53.6%) were found in the Africa, Western Pacific (52.8%) and South-East Asia regions (51.3%), respectively. The lowest proportion of undiagnosed diabetes was observed in North America and the Caribbean (24.2%).Diabetes surveillance needs to be strengthened to reduce the prevalence of UDM, particularly in low- and middle-income countries. The prevalence of diabetes is increasing worldwide. The International Diabetes Federation (IDF) estimates that 536.6 million people are living with diabetes (diagnosed or undiagnosed) in 2021, and this number is projected to increase by 46%, reaching 783.2 million by 2045 [34]. As previous IDF estimates and other studies have shown, approximately 50% of all individuals with diabetes are unaware of their condition [1], [2], [3]. From a clinical perspective, earlier identification during the asymptomatic stage is important to permit earlier initiation of treatment to prevent or delay the development of micro- and macrovascular complications. Studies have shown that a person may spend 5–6 years in an asymptomatic phase of pre-diabetes and type 2 diabetes mellitus (T2DM) before being diagnosed [4], during which time micro- and macrovascular complications may develop [5]. In 2013, it was estimated that 60% of people with T2DM were asymptomatic at the time of diagnosis [6]. A large cross-sectional study in Denmark found that 35% of previously undiagnosed participants had complications at the time of diagnosis, of whom 12% had microvascular, 17% had macrovascular, and 6% showed both micro- and macrovascular complications [5]. Of those with microvascular complications, retinopathy was the most frequent (13%), followed by neuropathy (4%) and nephropathy (3%). Ischemic heart disease was the most common macrovascular complication noted (15%), followed by atherosclerotic cerebrovascular (5%) and peripheral arterial disease (2%). However, since Denmark has a highly developed health system with a strong capacity for screening and diagnosis, these data may not be generalisable to countries lacking similar sophisticated health systems [7], [8], [9]. Indeed, without the same established medical infrastructure and resources necessary for early detection, middle- and low-income countries in other regions of the world may be characterised by higher numbers of people diagnosed with diabetes after the onset of complications. Maintaining good glycaemic control is the mainstay for the prevention of diabetes complications [10], [11], [12]. Thus, better systems for screening which identify individuals with undiagnosed diabetes earlier in the disease process – thus allowing the opportunity for treatment – will reduce morbidity and mortality associated with this disease. Accurate measurement of the burden of undiagnosed diabetes (UDM) is also critical to monitor public health efforts related to diabetes screening and diagnosis. The IDF first produced estimates of UDM in 2011 [13], providing a global-scale quantification of this burden. Such estimates are important as they provide the necessary data for governments and health care systems to benchmark diabetes prevention activities.

# Automated microaneurysms detection for early diagnosis of diabetic retinopathy: A Comprehensive review

[Diabetic retinopathy](https://www.sciencedirect.com/topics/medicine-and-dentistry/diabetic-retinopathy) (DR), a chronic disease in which the retina is damaged due to small vessel damage caused by diabetes mellitus, is one of the leading causes of [vision impairment](https://www.sciencedirect.com/topics/medicine-and-dentistry/visual-impairment) in diabetic patients. Detection of the earliest clinical sign of the advent of DR is a critical requirement for intervention and effective [treatment](https://www.sciencedirect.com/topics/medicine-and-dentistry/therapeutic-procedure). Ophthalmologists are trained to identify DR, based on examining specific minute changes in the eye - [microaneurysms](https://www.sciencedirect.com/topics/medicine-and-dentistry/microaneurysm), [retinal haemorrhages](https://www.sciencedirect.com/topics/medicine-and-dentistry/retinal-haemorrhage), [macular edema](https://www.sciencedirect.com/topics/medicine-and-dentistry/macular-edema) and changes in the [retinal blood vessels](https://www.sciencedirect.com/topics/medicine-and-dentistry/retinal-blood-vessel). Segmentation of microaneurysms (MA) is a critical requirement for the early diagnosis of DR and has been the primary focus of the research community over the past few years. In this work, a [systematic review](https://www.sciencedirect.com/topics/medicine-and-dentistry/systematic-review) of existing literature is carried out to examine the diagnostic use of automated MA detection and segmentation for early DR diagnosis. We mainly focus on existing early DR diagnosis techniques to understand their strengths and weaknesses. Though early diagnosis is performed using colour [fundus photography](https://www.sciencedirect.com/topics/medicine-and-dentistry/fundus-photography), [fluorescein angiography](https://www.sciencedirect.com/topics/medicine-and-dentistry/fluorescein-angiography) or [optical coherence tomography angiography](https://www.sciencedirect.com/topics/medicine-and-dentistry/optical-coherence-tomography-angiography) images, our study is limited to colour fundus based techniques. The early DR diagnosis methodologies reviewed in this article can be broadly classified into classical image processing, conventional machine learning (ML), and deep learning (DL) based techniques. Though significant progress has been achieved in these three classes of early DR diagnosis, several challenges and gaps still exist, underscoring a considerable scope for the development of fully automated, user-friendly early DR diagnosis and grading systems. We discuss in detail the challenges that need to be addressed in designing such effective, efficient, and robust algorithms for early DR diagnosis systems and also the ample scope for future research in this area. Diabetes Mellitus (DM) is a chronic disease characterized by increased blood sugar levels. Diabetic patients often suffer from severe damage to the blood vessels and nerves, thereby affecting the eyes, heart, kidneys and the brain. According to global reports on diabetes released by World Health Organization (WHO) in 2016 [[1]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0001), the number of adults who have diabetes has nearly doubled in just 25 years - from 108 million in 1980 to 422 million in 2014. In India, 102–258 million people are suffering from DM as of 2016 [[1]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0001). One of the critical complications that DM patients suffer from is [Diabetic retinopathy](https://www.sciencedirect.com/topics/medicine-and-dentistry/diabetic-retinopathy) (DR), characterized by damage to the blood vessels of the light-sensitive tissue at the back of the eye (retina) through fluid build-up. About 30% of the DM population has DR, and if left untreated, can develop a more severe form called [Proliferative DR](https://www.sciencedirect.com/topics/medicine-and-dentistry/proliferative-diabetic-retinopathy) (PDR), leading to severe [vision loss](https://www.sciencedirect.com/topics/medicine-and-dentistry/visual-impairment). DR occurs in predictable progression with minor variations in the order of their appearance and is the leading cause of legal vision loss among working-age individuals suffering from DM. As per the American Academy of [Ophthalmology](https://www.sciencedirect.com/topics/medicine-and-dentistry/ophthalmology) (AAO), DR is mainly classified into non-proliferative DR (NPDR) and proliferative DR (PDR) based on the retinal signs. As per Early [Treatment](https://www.sciencedirect.com/topics/medicine-and-dentistry/therapeutic-procedure) of DR Study (ETDRS)[1](https://www.sciencedirect.com/science/article/pii/S2666990021000124#fn0001), [microaneurysms](https://www.sciencedirect.com/topics/medicine-and-dentistry/microaneurysm) are the only clinical sign that is indicative of the onset of mild NPDR. During the initial stages of NPDR, retinal capillaries are damaged due to [Hyperglycemia](https://www.sciencedirect.com/topics/medicine-and-dentistry/hyperglycemia), which weakens the [capillary walls](https://www.sciencedirect.com/topics/medicine-and-dentistry/capillary-wall) and results in microaneurysms (MA) [[2]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0002). MA are the small outpouchings of the [retinal vessel](https://www.sciencedirect.com/topics/medicine-and-dentistry/retinal-blood-vessel) lumens, which eventually rupture to form haemorrhages (H) deep within the retina. Because of their appearance on [retinal examination](https://www.sciencedirect.com/topics/medicine-and-dentistry/retinal-examination), they are called ’dot’ and ’blot’ haemorrhages. Weakened vessels allow leakage of clear fluid called [transudates](https://www.sciencedirect.com/topics/medicine-and-dentistry/transudate) and a lipoprotein-rich fluid called exudates into the retina. Fluid deposition in the macula, called [Macular Edema](https://www.sciencedirect.com/topics/medicine-and-dentistry/macular-edema), hinders its normal function. If diagnosed early, DR associated complications and vision loss can be prevented or controlled due to the availability of advanced treatment modalities. Proliferative DR (PDR) is a [complication of diabetes](https://www.sciencedirect.com/topics/medicine-and-dentistry/complication-of-diabetes) caused by a continued restriction in [blood supply](https://www.sciencedirect.com/topics/medicine-and-dentistry/vascularity) to tissues, causing a shortage of oxygen that is needed for cellular metabolism (Ischemia) [[3]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0003). The retinal cells respond by releasing [angiogenic factors](https://www.sciencedirect.com/topics/medicine-and-dentistry/angiogenic-factor) like vascular endothelial growth factor (VEGF) to cope with the retinal high metabolic requirement. To bypass the damaged vessels, VEGF stimulates the growth of new leaky, fragile, and often misdirected retinal blood vessels called neovascularization. The new vessels not only grow into the retina but also into the vitreous. While the vitreous shrinks, it pulls on these fragile vessels and can tear, resulting in [vitreous haemorrhages](https://www.sciencedirect.com/topics/medicine-and-dentistry/vitreous-hemorrhage) and sudden vision impairment. Screening for DR is crucial since most patients do not experience any symptoms until the advanced stages of the disease, thereby reducing the efficacy of the treatments. [Table 1](https://www.sciencedirect.com/science/article/pii/S2666990021000124#tbl0001) presents more information on the classification of DR based on the observed clinical findings of the retina [[4]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0004). The accurate detection of microaneurysms (MA) and dot haemorrhages is a crucial step for early detection of DR as these are typically the earliest clinically recognizable signs [[5]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0005). These two often have a similar fundoscopic appearance and require [Fluorescein Angiography](https://www.sciencedirect.com/topics/medicine-and-dentistry/fluorescein-angiography) to differentiate between them. However, both types of lesions could be considered together in grading DR. It has been reported that MA’s identification plays a significant role in identifying the stage of NPDR (shown in [Table 2](https://www.sciencedirect.com/science/article/pii/S2666990021000124#tbl0002)). However, manual detection of MA in fundus [photographs](https://www.sciencedirect.com/topics/medicine-and-dentistry/photograph) is a time-consuming task, as their pixels are similar in appearance to that of surrounding blood vessels. It is also hard to distinguish them from background variations as they are typically low in contrast. Though several published reports discuss the existing body of work in the development of computer-based DR screening using [retinal fundus](https://www.sciencedirect.com/topics/medicine-and-dentistry/fundus-eye) images [[6]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0006), [[7]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0007), [[8]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0008), [[9]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0009), [[10]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0010), [[11]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0011), [[12]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0012), [[13]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0013), [[14]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0014), [[15]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0015), [[16]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0016), a detailed study on early DR screening approaches and the associated challenges are not explored. Furthermore, the existing surveys mainly focus on deep learning-based methods only. A comprehensive review of major works addressing the problem of early DR screening by automated detection/segmentation of microaneurysms[2](https://www.sciencedirect.com/science/article/pii/S2666990021000124#fn0002) is presented in this paper. The publicly available datasets, preprocessing methods employed, and the quantitative comparison techniques using standard performance metrics are also discussed in detail. The challenges in the process are also emphasized, and potential solutions highlighting future research directions for the early DR diagnostic systems are presented. The PRISMA guidelines [[17]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0017) are followed for identifying the relevant prior work. Multiple databases, viz. Scopus, PubMed, IEEE Xplore, ACM, Springer and arXiv are reviewed exhaustively. The search terms include (((early AND (diagnosis OR detection OR prediction) AND (diabetic retinopathy)) OR ((microaneurysms OR microaneurysm) AND (segmentation OR detection)) along with synonyms on the titles, keywords and abstracts. Duplicates and irrelevant articles are removed, and further filtering is performed based on predefined inclusion/exclusion criteria. The methodology adopted for the [systematic review](https://www.sciencedirect.com/topics/medicine-and-dentistry/systematic-review) of works on early DR diagnosis is presented in [Fig. 1](https://www.sciencedirect.com/science/article/pii/S2666990021000124#fig0001). The search was performed to identify all studies in which [retinal imaging](https://www.sciencedirect.com/topics/medicine-and-dentistry/retinal-imaging) is used for the early diagnosis of DR, including older and recent reviews, since 1983. Only English language articles published in peer-reviewed journals and core conferences are considered. Also, only original studies are considered, excluding the case reports and previous review/survey articles. The articles that do not provide validation of their works are also excluded. Research works that used diagnostic images other than colour fundus images are included (8 papers), but a detailed discussion is not provided due to the nonavailability of results validated with publicly available data.

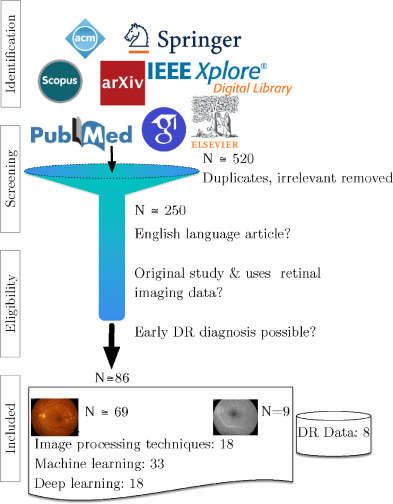
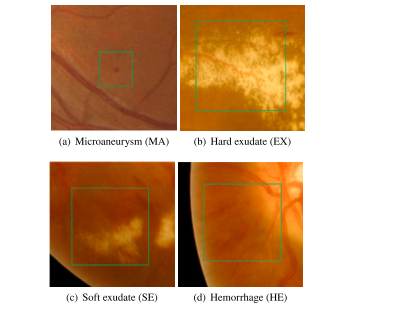
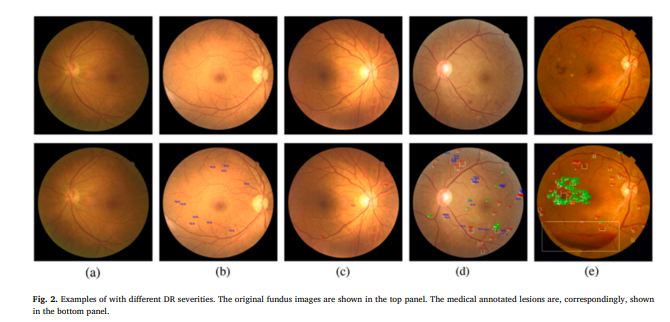


Fig. 1. [Systematic review](https://www.sciencedirect.com/topics/medicine-and-dentistry/systematic-review) process as per PRISMA guidelines [[17]](https://www.sciencedirect.com/science/article/pii/S2666990021000124#bib0017).

Based on these criteria, the resultant initial cohort included around 250 articles, of which 86 met the selection criteria. Of these, 8 provided the baseline results along with the labelled dataset, resulting in 69 articles being retained for this study. Based on their implementation techniques, the cohort is broadly classified into three categories - classical image processing-based works that use hand-crafted filters and thresholding methods (18 articles), conventional ML-based techniques that use hand-crafted features (33 articles), and end-to-end, data-driven DL based methods (18 articles). This work is a comprehensive review of the current literature, with an emphasis on quantitative and qualitative comparisons. In addition, persistent problems and scientific research gaps in early DR diagnosis are thoroughly explored.

**Fine-grained attention & knowledge-based collaborative network for diabetic retinopathy grading** Accurate diabetic retinopathy (DR) grading is crucial for making the proper treatment plan to reduce the damage caused by vision loss. This task is challenging due to the fact that the DR related lesions are often small and subtle in visual differences and intra-class variations. Moreover, relationships between the lesions and the DR levels are complicated. Although many deep learning (DL) DR grading systems have been developed with some success, there are still rooms for grading accuracy improvement. A common issue is that not much medical knowledge was used in these DL DR grading systems. As a result, the grading results are not properly interpreted by ophthalmologists, thus hinder the potential for practical applications. This paper proposes a novel fine-grained attention & knowledge-based collaborative network (FA+KC-Net) to address this concern. The fine-grained attention network dynamically divides the extracted feature maps into smaller patches and effectively captures small image features that are meaningful in the sense of its training from large amount of retinopathy fundus images. The knowledge-based collaborative network extracts a-priori medical knowledge features, i.e., lesions such as the microaneurysms (MAs), soft exudates (SEs), hard exudates (EXs), and hemorrhages (HEs). Finally, decision rules are developed to fuse the DR grading results from the fine-grained network and the knowledgebased collaborative network to make the final grading. Extensive experiments are carried out on four widely-used datasets, the DDR, Messidor, APTOS, and EyePACS to evaluate the efficacy of our method and compare with other state-of-the-art (SOTA) DL models. Simulation results show that proposed FA+KC-Net is accurate and stable, achieves the best performances on the DDR, Messidor, and APTOS datasets. Diabetic retinopathy (DR) is a prevalent long-term complication that often arises in individuals with diabetes. It stands as the predominant factor affecting blindness in the adult working-age population worldwide [1] [2]. Nearly one-third of individuals diagnosed with diabetes develop diabetic retinopathy, indicating that every diabetic patient is susceptible to this condition [3]. The most common abnormalities of DR are the MA, EX, SE, and HE. Fig. 1(a-d) illustrates these diseases. The DR can be categorized

****

****

into five categories: no DR 0, mild DR 1, moderate DR 2, severe DR 3 and proliferative DR 4. Fig. 2(a-e) shows the DR categories [4]. Timely detection of diabetic retinopathy with accurate grading can effectively slow down or prevent the progression of visual impairment [5] [6]. Often, doctors pay more attention to the cases of moderate DR 2 and severe DR 3, because they are very difficult to be correctly distinguished. The effect is whether the patient needs immediate further examination and treatment. The conventional way of DR detection requires ophthalmologists to evaluate fundus images, which is subjective and timeconsuming [5].

Ophthalmologists usually identify and grade DR based on the types and numbers of related lesions. However, with more people developing diabetes [7] [8], demand for real-time analysis of fundus images increases dramatically. This brings heavy burden for the limited ophthalmologists. Therefore, it is crucial to develop a reliable automatic DR grading method to relieve the pressure of ophthalmologists and to get more timely treatment to the patients. The computer-aided method provides objective evaluation for ophthalmologists. The DR grading results may bring the attention to the doctors, who, depending on their levels of knowledge and expertise, may provide varying grades for the same patient due to the subtle intra-class variations [9]. Recently, DL method has achieved remarkable advancements in the domain of medical image, such as image classification [10], detection [11], and tumor segmentation [12]. CNN (convolutional neural network) architectures have found widespread application in the field of medical image processing, catering to diverse tasks and objectives. In the field of DR grading, many auto-grading models Heliyon 9 (2023) e17217 3 M. Tian, H. Wang, Y. Sun et al. have been presented with promising results [13] [14] [15]. Vo et al. [13] presented a structure combining kernels with multiple losses and VGG net with extra kernel to learn the features. Vives-Boix et al. [14] combined CNNs with synaptic metaplasticity to improve the convergence rate and performance. Bodapati et al. [15] proposed a composite neural network architecture that incorporates a gated-attention mechanism for the task. Despite the remarkable achievements of existing DR grading methods, there are still rooms for improvement on the DR grading performance, especially for the distinction between two adjacent severity levels.

A possible way to cope with this challenge is to extract sufficient relevant lesion information that is strongly associated with the severity of DR. Since some lesions only contain a few pixels, these lesions are easily neglected when multiple convolutions are used to extract features. Certain DR grading models employ modules similar to the spatial attention mechanisms or filters of multiple sizes to exploit detailed lesion information [13] [16]. These methods offer limited performance improvement because DR grading is essentially a fine-grained image classification task [17]. It aims to capture the discriminative details, however, the subtle intra-class variation and inter-class similarity make it extremely challenging [18]. Further, it would be more meaningful for practical applications if DR grading methods can provide satisfactory medical interpretation for ophthalmologists. Most methods do not effectively incorporate relevant medical knowledge. And the black-box behavior has become an obstacle to clinical application. This paper presents a novel FA+KC-Net model for DR grading.

The model contains a fine-grained attention network (FA-Net) and a knowledge-based collaborative network (KC-Net). The FA-Net is designed to effectively capture detailed small visual features and to suppress the redundant information. The KC-Net employs a lesion detector and applies specifically designed rules from medical knowledge to capture as many DR related lesions as possible. Final grading results are obtained by fusing the results from the FA-Net and KC-Net. A comprehensive set of experiments is conducted to validate and demonstrate the efficacy of the proposed method on four widely recognized public datasets: DDR, EyePACS, Messidor, and APTOS. Results obtained indicate that the FA+KC-Net outperforms other methods on the DDR, Messidor, and APTOS datasets. It is also found that the KC-Net offers an interesting benefit. Thorough DR grading evaluation on the DDR dataset reveals that a few cases might be labeled with the incorrect DR grading. Majority of the label errors occur for the cases being labeled as DR 3. Results by the KC-Net indicate that the lesions are not that severe and should be labeled as DR 1 or DR 2. Our findings are confirmed by five ophthalmologists from the West China Hospital, Sichuan University. Several significant contributions: 1. Design a fine-grained attention network (FA-Net) that dynamically divides the extracted feature maps into smaller patches and efficiently extracts distinctive features.

2. Propose a knowledge-based collaborative network (KC-Net) for DR grading by applying a-priori medical knowledge so as to provide the detected lesions with medical explanation. 3. Develop the decision rules to integrate the grading results from the KC-Net into the FA-Net for optimal DR grading results.

# Robust Vessel Segmentation in Fundus Images

# One of the most common modalities to examine the human eye is the eye-fundus photograph. The evaluation of fundus photographs is carried out by medical experts during time-consuming visual inspection. Our aim is to accelerate this process using computer aided diagnosis. As a first step, it is necessary to segment structures in the images for tissue differentiation. As the eye is the only organ, where the vasculature can be imaged in an in vivo and noninterventional way without using expensive scanners, the vessel tree is one of the most interesting and important structures to analyze. The quality and resolution of fundus images are rapidly increasing. Thus, segmentation methods need to be adapted to the new challenges of high resolutions. In this paper, we present a method to reduce calculation time, achieve high accuracy, and increase sensitivity compared to the original *Frangi* method. This method contains approaches to avoid potential problems like specular reflexes of thick vessels. The proposed method is evaluated using the *STARE* and *DRIVE* databases and we propose a new high resolution fundus database to compare it to the state-of-the-art algorithms. The results show an average accuracy above 94% and low computational needs. This outperforms state-of-the-art methods. In ophthalmology the most common way to examine the human eye is to take an eye-fundus photograph and to analyse it. During this kind of eye examinations a medical expert acquires a photo of the eye-background through the pupil with a fundus camera. The analysis of these images is commonly done by visual inspection. This process can require hours in front of a computer screen, in particular in case of medical screening. An example fundus image is shown in Figure [1](https://www.hindawi.com/journals/ijbi/2013/154860/fig1/).

# 

Our goal is to speed up the diagnosis by processing the images using computer algorithms to find and highlight the most important details. In addition we aim to automatically identify abnormalities and diseases with minimal human interaction. Due to the rapidly increasing spatial resolution of fundus images, the common image processing methods which were developed and tested using low resolution images have shown drawbacks in clinical use. For this purpose, a new generation of methods needs to be developed. These methods need to be able to operate on high resolution images with low computational complexity. In this paper, we would like to introduce a novel vessel segmentation method with low computational needs and a public available high resolution fundus database with manually generated gold standards for evaluation of retinal structure segmentation methods. The proposed algorithms include modifications to the method proposed by Frangi et al. [1] to decrease the running time and to segment specular reflexes of thick vessels, which are not visible in lower resolution fundus images.

# Human Visual System-Based Fundus Image Quality Assessment of Portable Fundus Camera Photographs Telemedicine and the medical “big data” era in ophthalmology highlight the use of non-mydriatic ocular fundus photography, which has given rise to indispensable applications of portable fundus cameras. However, in the case of portable fundus photography, non-mydriatic image quality is more vulnerable to distortions, such as uneven illumination, color distortion, blur, and low contrast. Such distortions are called generic quality distortions. This paper proposes an algorithm capable of selecting images of fair generic quality that would be especially useful to assist inexperienced individuals in collecting meaningful and interpretable data with consistency. The algorithm is based on three characteristics of the human visual system—multi-channel sensation, just noticeable blur, and the contrast sensitivity function to detect illumination and color distortion, blur, and low contrast distortion, respectively. A total of 536 retinal images, 280 from proprietary databases and 256 from public databases, were graded independently by one senior and two junior ophthalmologists, such that three partial measures of quality and generic overall quality were classified into two categories. Binary classification was implemented by the support vector machine and the decision tree, and receiver operating characteristic (ROC) curves were obtained and plotted to analyze the performance of the proposed algorithm. The experimental results revealed that the generic overall quality classification achieved a sensitivity of 87.45% at a specificity of 91.66%, with an area under the ROC curve of 0.9452, indicating the value of applying the algorithm, which is based on the human vision system, to assess the image quality of non-mydriatic photography, especially for low-cost ophthalmological telemedicine applications.

# OPHTHALMOLOGICAL early detection using retinal photography prevents both vision impairment and the consequences of untreated eye disease. However, the current

# lack of medical resources in some areas of developing countries, like China, leads to unfavorable outcomes. Non-mydriatic ocular fundus photography seems to be a promising solution, especially for retinal disease, when combined with telemedicine [1], [2], because it does not require pupil dilation and it can be done with a portable fundus camera. Portable digital fundus photography differs from traditional fundus photography because it is done with the camera fixed to operator's hands rather than being on a permanent fixture. However, such operating conditions can be vulnerable to problems with the quality of the digital retinal image, such as uneven luminance, fluctuations in focus, and patients' movements. Hence, evaluating the image quality of portable fundus camera imaging-systems is of great importance. The evaluation of fundus image quality involves a computer-aided retinal image analysis-system that is designed to assist ophthalmologists to detect eye diseases [3], such as age-related macular degeneration [4], glaucoma [5], [6], and diabetic retinopathy [7].

# The objective quality evaluation of fundus images, which plays a major role in automatically selecting diagnosis-accessible fundus images among the outputs of digital fundus photography, is a descendant of subjective quality evaluation. Subjective quality evaluation is performed by experienced ophthalmologists who grade the quality of fundus images by comparing differences in the images to be graded with excellent quality images, based on their prior knowledge of excellent image quality. Such prior knowledge is acquired either from the human visual system (HVS), which is a complex biological system [8], or from technical training in ophthalmic diagnosis. Based on their prior knowledge, ophthalmological experts can grade fundus image quality with confidence; however, their subjective quality evaluation is as laborious in practice as it is expensive and time-consuming. Objective fundus image quality assessment is aimed at providing ophthalmologists with an alternative grading procedure, which not only takes less time, but also has equivalent accuracy. Research related to the objective assessment of fundus image quality has been conducted for decades. The methods proposed in these studies can be classified into two major categories: generic feature-based methods [9]–[11] and structural feature-based methods [12]–[14].

# Generic feature-based methods deal with global distortions, such as uneven illumination, blurring effects from being out of focus, and low contrast. Lee and Wang [15] presented an explicit template that was mathematically approximated by a Gaussian model to extract images of desired quality from a set of images. Convolution of the template with the intensity histogram of a retinal image was computed as generic quality. Fasih et al. [11] developed. a generic retinal image quality estimation system that employed just noticeable blur (JNB), an HVS characteristic [16], combined with texture features [17]. Instead of finding an explicit template, implicit templates were utilized later on [10], including machine learning techniques such as the distance threshold of the k-Nearest Neighbours Classifier [18], support vectors embedded in a Support Vector Machine (SVM) [19], or weights in a Neural Network [20]. The HVS has many characteristics described by mathematics that humans can recognize as various patterns, such as colour [21], orientation [22], contour [23], motion [24], and frequency variation [25]. When applied to retinal image quality assessment, low-level characteristics of the HVS can extract generic features, such as illumination and colour, while high-level characteristics can extract structural features, such as vessel edges and macular texture. Here, we apply low-level HVS characteristics to generic quality assessment and propose an integrated HVS-based generic quality assessment algorithm as a starting point. Generic quality involves three parameters: illumination and colour, focus, and contrast.

# Three low-level characteristics of the HVS, including multi-channel sensation, just noticeable blur, the contrast sensitivity function, were employed to evaluate the three parameters, respectively. The rest of the paper is organized as follows: Section II introduces materials containing public and proprietary retinal datasets and the subjective evaluation of images in these datasets; Section III describes the proposed algorithm, which consists of three parts: image preprocessing, HVS-based feature extraction, and a machine learning procedure; Section IV presents the results of tests of the algorithm on proprietary and public datasets, and compares these results with ophthalmologists' subjective evaluations; and Sections V and VI present the discussion and conclusion, respectively.

# EfficientNetV2 Based Ensemble Model for Quality Estimation of Diabetic Retinopathy Images from DeepDRiD

# Diabetic retinopathy (DR) is one of the major complications caused by diabetes and is usually identified from retinal fundus images. Screening of DR from digital fundus images could be time-consuming and error-prone for ophthalmologists. For efficient DR screening, good quality of the fundus image is essential and thereby reduces diagnostic errors. Hence, in this work, an automated method for quality estimation (QE) of digital fundus images using an ensemble of recent state-of-the-art EfficientNetV2 deep neural network models is proposed. The ensemble method was cross-validated and tested on one of the largest openly available datasets, the Deep Diabetic Retinopathy Image Dataset (DeepDRiD). We obtained a test accuracy of 75% for the QE, outperforming the existing methods on the DeepDRiD. Hence, the proposed ensemble method may be a potential tool for automated QE of fundus images and could be handy to ophthalmologists.

Diabetic retinopathy (DR) is a common disease caused by diabetes, majorly affecting working individuals and leading to loss of vision. By 2040, it is estimated that 600 million people will suffer from diabetes, and approximately one third of them will have a chance of getting DR [[1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B1-diagnostics-13-00622)]. An ophthalmologist usually identifies DR by visual examination of digital fundus images for the presence of one or more retinal lesions such as microaneurysms, soft exudates, hemorrhages, and hard exudates [[2](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B2-diagnostics-13-00622)]. DR can broadly be classified into non-proliferative DR (NPDR) and proliferative DR (PDR). The preliminary stage of DR is NPDR, where the microaneurysms are visible in the digital fundus image, and the advanced stage of DR is PDR which can lead to severe vision loss. The NPDR is further subdivided into three types: mild, moderate, and severe NPDR. The international clinical DR severity scale contains five grades to classify fundus images—grade 0 is no apparent retinopathy, grade five is PDR, and the types mentioned above of NPDR are classified as grade one, two, and three, respectively.

The manual evaluation of fundus images may create a severe burden on ophthalmologists. Moreover, accurate grading of DR requires trained healthcare professionals and manual grading could be prone to errors while handling large amounts of data. Hence, automated methods for DR screening are warranted to reduce diagnostic oversights by ophthalmologists and healthcare practitioners. Furthermore, poor-quality digital fundus images due to uneven illumination, blurring, and other artifacts can lead to false positives. Hence, it is vital to first estimate the quality of acquired funds images before proceeding with DR grading [[3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B3-diagnostics-13-00622)]. Therefore, fully automated methods for accurate quality estimation (QE) of digital fundus images are in demand since the ratio of doctors to patients is deteriorating. Overall, there is a need for objective evaluation of fundus image quality to mimic the quality diagnosis of ophthalmologists.

In the past decade, several state-of-the-art deep learning (DL) architectures, including AlexNet [[4](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B4-diagnostics-13-00622)], VGGs [[5](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B5-diagnostics-13-00622)], GoogLeNet [[6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B6-diagnostics-13-00622)], ResNet [[7](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B7-diagnostics-13-00622)], DenseNet [[8](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B8-diagnostics-13-00622)], EfficientNets [[9](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B9-diagnostics-13-00622),[10](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B10-diagnostics-13-00622)], and, recently, vision transformer (ViT) [[11](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B11-diagnostics-13-00622)] based models were developed for various computer vision tasks such as object localization, object detection, and classification. Even though training large DL models from scratch requires massive data, transfer learning (TL) could facilitate adapting these already trained models for new classification tasks, thus eliminating the need for huge data for retraining. Furthermore, both TL and DL have been playing a major role in healthcare by building automated diagnostic systems for several diseases using medical images from radiographs, computed tomography, digital fundus images, positron emission tomography, and magnetic resonance imaging, etc.

These systems are primarily used for diagnostic and prognostic tasks and also assist medical practitioners in several scenarios such as faster data acquisition and quality control [[12](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B12-diagnostics-13-00622),[13](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B13-diagnostics-13-00622),[14](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B14-diagnostics-13-00622)]. EfficientNetV2 is one of the recently developed DL architectures based on progressive learning with a combination of training-aware neural architecture search and compound scaling to improve both the training speed and parameter efficiency [[9](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B9-diagnostics-13-00622)], and it outperformed several previous state-of-the-art models including ViTs in image classification tasks on the ImageNet challenge. Therefore, the following are the contributions of this work:

* A fully automated method for the overall QE of digital fundus images is proposed using an ensemble of pretrained EfficientNetV2- small (S), medium (M), and large (L) models since model ensembling was effective in some previous studies [[15](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B15-diagnostics-13-00622),[16](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B16-diagnostics-13-00622)].
* The proposed ensemble model is cross-validated and tested on a large publicly available dataset called the Deep Diabetic Retinopathy Image Dataset (DeepDRiD), as the QE of fundus images from this dataset seems challenging [[3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955381/#B3-diagnostics-13-00622)].
* The ability of the proposed ensemble model for overall QE is further stratified concerning DR disease severity.

# Deep learning for quality assessment of retinal OCT images

# Optical coherence tomography (OCT) is a promising high-speed, non-invasive imaging modality providing high-resolution retinal scans. However, a variety of external factors such as light occlusion and patient movement can seriously degrade OCT image quality, which complicates manual retinopathy detection and computer-aided diagnosis. As such, this study first presents an OCT image quality assessment (OCT-IQA) system, capable of automatic classification based on signal completeness, location, and effectiveness. Four CNN architectures (VGG-16, Inception-V3, ResNet-18, and ResNet-50) from the ImageNet classification task were used to train the proposed OCT-IQA system via transfer learning. The ResNet-50 with the best performance was then integrated into the final OCT-IQA network. The usefulness of this approach was evaluated using retinopathy detection results. A retinopathy classification network was first trained by fine-tuning Inception-V3 model. The model was then applied to two test datasets, created randomly from the original dataset, one of which was screened by the OCT-IQA system and only included high quality images while the other was mixed by high and low quality images. Results showed that retinopathy detection accuracy and area under curve (AUC) were 3.75% and 1.56% higher, respectively, for the filtered data (compared with the unfiltered data). These experimental results demonstrate the effectiveness of the proposed OCT-IQA system and suggest that deep learning could be applied to the design of computer-aided systems (CADSs) for automatic retinopathy detection.

Optical coherence tomography (OCT), which can image retinal structures in vivo [[1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r1)], has been widely applied in diagnostic ophthalmology due to its ease-of-use, lack of ionizing radiation, and high resolution [[2](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r2)]. There are approximately 30 million OCT procedures performed worldwide each year [[3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r3)], with hundreds of consecutive B-scans comprising the majority of each procedure. This produces large quantities of data and limits the manual evaluation of individual images. Recent developments in computer-aided diagnostic systems (CADSs) have aided in retinopathy diagnosis and reduced the workload for clinicians. This has reduced processing times by accelerating image evaluation and improving diagnoses [[4](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r4)].

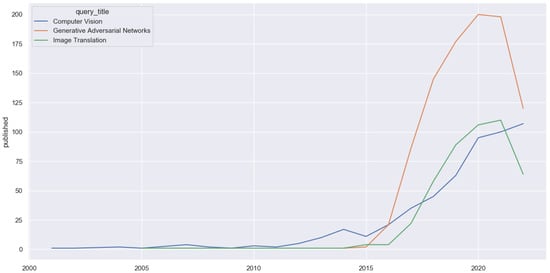
Previous studies have focused on screening retinal diseases using OCT images and an automatic CAD system [[5](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r5)–[9](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r9)]. Farsiu et al. assessed OCT scan quality by comparing results with manual evaluations performed by experts, excluding poor quality images from the dataset [[5](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r5)]. Wang et al. eliminated poor quality images during pre-processing and Kermany et al. removed half of the original OCT data in an initial quality review, which decreased the sample size [[6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r6),[9](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r9)]. Rasti et al. and Treder et al. used only images from public datasets that passed an image quality assessment (IQA) process. As is common, none of these studies used low-quality images in training the CADS. However, in practice, low quality retinal OCT images are inevitably produced due to insufficient contrast, lighting conditions, patient movement, and signal occlusion, which can cause signal void or blurring and prevent images from being suitable for diagnosis. As such, the use of an IQA in a CADS is critical for eliminating low quality images.

IQA methods can be divided into subjective and objective assessments. Subjective IQA methods were applied in the studies listed above, wherein experts manually eliminated unsatisfactory images [[5](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r5),[6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r6),[9](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r9)]. Objective IQA methods can assist clinicians by automatically evaluating image quality, without requiring manual intervention. Several objective IQA algorithms have been proposed in recent years, including reference and blind IQA techniques. Reference methods primarily involve determining image parameters. For example, Ishikawa et al. [[10](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r10)] calculated the signal-to-noise ratio (SNR) and an image quality (IQ) metric to assess OCT scans. Stein et al. [[11](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r11)] introduced the quality index (QI), a new IQA parameter based on histogram information. Liu et al. [[12](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r12)] proposed signal deviation (SD), which considers the standard deviation of measured intensities, and Huang et al. [[13](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r13)] used a maximum tissue contrast index (mTCI) to quantify the image signal from multiple devices. Reference IQA can automatically and objectively eliminate unqualified images but robustness and accuracy suffer from a heavy dependence on image parameters (i.e., intensity, SNR, and signal strength). This can amplify parameter calculation errors in subsequent assessments. On the other hand, these image parameters mentioned above often fail to represent image quality. For example, an image with high SNR could be unacceptable for retinopathy diagnosis because of an off-centered signal, making the reference IQA unsuitable.

Blind IQA has become more generalizable with the development of machine learning and does not require additional information beyond the original data [[14](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r14)]. As such, it is commonly used for fundus images classification [[15](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r15)–[21](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r21)] and has proven to be effective in eliminating low-quality images. Few studies have investigated the use of blind IQA methods based on OCT until 2019. Kauer et al. proposed an AQuANet to classify OCT B-scans into ‘good,’ ‘bad,’ ‘upper,’ and ‘lower’ categories using A-scans [[22](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r22)]. However, this study only investigated the position of the signal. Zhang et al. proposed a blind IQA architecture based on ResNet50. In this process, OCT images were partitioned into non-overlapping patches for preprocessing. ResNet50 was then adopted as a feature extractor and support vector regression was used to train the IQA model. However, this study primarily focused on assessing signal intensity [[23](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6913385/#r23)]. These two blind OCT-IQA techniques neglected both signal effectiveness and completeness, which are critical for retinopathy detection.

In this study, we first develop a multi-class deep neural network for pre-filtering a retinopathy detection CADS. This approach pays attention to signal location, effectiveness and completeness, which can automatically assess the quality of retinal OCT images, including signal occlusion, signal centering, and the position of the region of interest. Deep neural networks are proficient in extracting image features, which become increasingly abstract as the layer depth increases. However, deep neural networks characteristically generate multiple parameters and require large quantities of data to avoid over-fitting during the training process. As such, we adopted transfer learning to develop an OCT-IQA network and avoid high computational costs. Four pre-trained CNN architectures (VGG, ResNet-18, ResNet-50, and Inception-V3) were implemented in the study, each of which won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). We fine-tuned the learned networks for our specific classification task and adopted the highest performing ResNet-50. In addition, a separate retinopathy detection model was developed by fine-tuning the Inception-V3 to test the influence of unqualified images on further CADS and demonstrate the necessity of IQA. Two test datasets collected from the original dataset were adopted for retinopathy detection, one of which was first fed to the OCT-IQA network to eliminate unqualified images and produce a ‘pure’ dataset. The other was composed of acceptable and unacceptable images with a 1:1 ratio, and labeled the ‘mixed’ set. Test results showed that retinopathy detection accuracy and area under curve (AUC) were 3.75% and 1.56% higher, respectively, for the pure data (compared with the mixed data), which demonstrated that image quality is a vital element in automatic retinopathy detection.

**UNSUPERVISED IMAGE-TO-IMAGE TRANSLATION: A REVIEW** Supervised image-to-image translation has been proven to generate realistic images with sharp details and to have good quantitative performance. Such methods are trained on a paired dataset, where an image from the source domain already has a corresponding translated image in the target domain. However, this paired dataset requirement imposes a huge practical constraint, requires domain knowledge or is even impossible to obtain in certain cases. Due to these problems, unsupervised image-to-image translation has been proposed, which does not require domain expertise and can take advantage of a large unlabeled dataset. Although such models perform well, they are hard to train due to the major constraints induced in their loss functions, which make training unstable. Since CycleGAN has been released, numerous methods have been proposed which try to address various problems from different perspectives. In this review, we firstly describe the general image-to-image translation framework and discuss the datasets and metrics involved in the topic. Furthermore, we revise the current state-of-the-art with a classification of existing works. This part is followed by a small quantitative evaluation, for which results were taken from papers. Image-to-image (I2I) translation aims to transfer an image from one domain to another while preserving the content of the given image. For example, we can take the famous horse to zebra translation, where the aim is to translate a horse image into a zebra image. The domains of zebra and horse images differ by some of characteristics given by the two sets of images (for example, zebras have stripes, and horses have bigger tails than zebras.) As shown in [**Figure 1**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#fig_body_display_sensors-22-08540-f001), the number of publications in image translation increased when generative adversarial networks (GANs) were proposed. In 2016 Isola et al. presented Pix2Pix [[**1**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B1-sensors-22-08540)], a conditional model that is able to translate an image from one domain to another using paired training. This work was followed by Pix2PixHD [[**2**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B2-sensors-22-08540)], which was translation of high-resolution images. However, even though these methods performed well when they were significant advances over the state-of-the-art, one major problem of these paired image-to-image translation methods is the paired dataset, a dataset where an image already has its translated counterpart. These paired datasets are hard, expensive or even impossible to obtain due to the parity constraint between the input and output.



**Figure 1.** Number of papers published per year in deep learning, generative adversarial networks, image-translation and image-to-image translation. Papers from the ArXiV database were counted.

Consequently, the research community has explored approaches to overcoming the need for paired datasets. For example, Bousmalis et al. [[**3**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B3-sensors-22-08540)] tried an unsupervised image-to-image translation method with a domain adaptation on the pixel space. With a more probabilistic point of view, Liu et al. [[**4**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B4-sensors-22-08540)] proposed an image translation method based on the shared latent space assumption. In the same way, Taigman et al. [[**5**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B5-sensors-22-08540)] proposed another architecture composed of a GAN and an input function *f*. The GAN is trained to generate wanted images and *f* converts an image to a latent representation, which is used by the GAN for the generation.

In 2017, Zhu et al. introduced CycleGAN [[**6**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B6-sensors-22-08540)], a network able to translate an image from one domain to another using cycle consistency. Nevertheless, despite this method breaking the paired constraint, it requires a lot of training iterations and suffers from instability. Since then, many applications have emerged (e.g., semantic to real [[**1**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B1-sensors-22-08540),[**6**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B6-sensors-22-08540),[**7**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B7-sensors-22-08540)], maps to satellites [[**1**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B1-sensors-22-08540),[**8**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B8-sensors-22-08540)] and satellites to street view [[**9**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B9-sensors-22-08540)]).

There remain challenges, the most prominent being:

**Complex Translations:** Translations that need heavy modifications to match the target space, such as geometrical transformations; or diverse images, such as those of landscapes.

**Effective Training:** In the context of unsupervised translation, cycle consistency is often used. However, it can be too restrictive in some cases, such as glass removal tasks [[**10**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B10-sensors-22-08540)]. This case forces the model to hide information in the translation in order to allow the backward translation to satisfy the cycle consistency [[**11**](https://www.mdpi.com/1424-8220/22/21/8540?type=check_update&version=1#B11-sensors-22-08540)]. Since that time, some implementations try to do without cycle consistency by proposing alternative losses. On the other hand, other approaches try to propose new discriminator architectures to stabilize the training.

**Data Scarcity:** Sometimes data are missing or are particularly difficult to collect. Some works have tried to reduce the data requirements of translation models by augmentation methods or directly with architecture changes.

This paper is aims to give an overview of principal works on unsupervised image-to-image translation (UI2IT), along with the current challenges and limitations. To accomplish these ends, this review is organized as follows: In the first place, we detail the general process of UI2IT. Secondly, we give an overview of datasets and metrics used for UI2IT, followed by the classification of methods. Thereafter, the main parts of this review are presented, namely, architecture changes, complex translation, data issues, attribute editing, guidance, disentanglement learning and contrastive learning methods. A short discussion about method comparison and current challenges is presented afterward, followed by the conclusions of the article.

# Dynamic and Real-Time Object Detection Based on Deep Learning for Home Service Robots

# Home service robots operating indoors, such as inside houses and offices, require the real-time and accurate identification and location of target objects to perform service tasks efficiently. However, images captured by visual sensors while in motion states usually contain varying degrees of blurriness, presenting a significant challenge for object detection. In particular, daily life scenes contain small objects like fruits and tableware, which are often occluded, further complicating object recognition and positioning. A dynamic and real-time object detection algorithm is proposed for home service robots. This is composed of an image deblurring algorithm and an object detection algorithm. To improve the clarity of motion-blurred images, the DA-Multi-DCGAN algorithm is proposed. It comprises an embedded dynamic adjustment mechanism and a multimodal multiscale fusion structure based on robot motion and surrounding environmental information, enabling the deblurring processing of images that are captured under different motion states. Compared with DeblurGAN, DA-Multi-DCGAN had a 5.07 improvement in Peak Signal-to-Noise Ratio (PSNR) and a 0.022 improvement in Structural Similarity (SSIM). An AT-LI-YOLO method is proposed for small and occluded object detection. Based on depthwise separable convolution, this method highlights key areas and integrates salient features by embedding the attention module in the AT-Resblock to improve the sensitivity and detection precision of small objects and partially occluded objects. It also employs a lightweight network unit Lightblock to reduce the network’s parameters and computational complexity, which improves its computational efficiency. Compared with YOLOv3, the mean average precision (mAP) of AT-LI-YOLO increased by 3.19%, and the detection precision of small objects, such as apples and oranges and partially occluded objects, increased by 19.12% and 29.52%, respectively. Moreover, the model inference efficiency had a 7 ms reduction in processing time. Based on the typical home activities of older people and children, the dataset Grasp-17 was established for the training and testing of the proposed method. Using the TensorRT neural network inference engine of the developed service robot prototype, the proposed dynamic and real-time object detection algorithm required 29 ms, which meets the real-time requirement of smooth vision.

With the development of the economy, improvement in living standards, and intensification of the aging trend of the global society, home service robots are witnessing increasing market demand. Although the most popular robots presently working in houses execute floor-cleaning tasks, home service robots are expected to do more. In particular, they are expected to play a significant role in assisting older people and children in modern families. To perform tasks such as completing housework and interacting with family members, these robots must be able to identify and locate their target objects in real-time [[1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B1-sensors-23-09482)], such as clothes in dressing tasks [[2](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B2-sensors-23-09482)] and door handles in door-opening tasks [[3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B3-sensors-23-09482)]. In addition, service robots usually need to change their position to facilitate task execution, which results in a high probability of blurring in the images collected by the camera on these robots. Blurred images cause difficulties in subsequent object detection processing. Moreover, objects that need to be handled in a living environment include a large number of small items (such as fruits, tableware, and toys).

Also, there always exists visual occlusions between objects. Furthermore, mobile robots are powered by batteries. So, they typically use small, embedded processors with limited power consumption, memory, and computing resources, which raises the problems of memory shortage and insufficient computing power. In summary, there are several prominent challenges in the development of object detection for home service robots, including motion blur [[4](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B4-sensors-23-09482),[5](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B5-sensors-23-09482)], small objects and locally occluded objects [[6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B6-sensors-23-09482)], and limited computing resources [[7](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B7-sensors-23-09482)].

Existing commercial off-the-shelf (COTS) solutions, such as global shutter and increasing frame rate, can reduce motion blur to some extent. However, in low-light conditions, they may cause the images collected to be too dark or to have excessive noise. Moreover, in high-contrast scenes, they may result in the loss of details in the bright and dark areas of the images. These issues make these solutions not entirely suitable for home service robot applications. In contrast, deblurring techniques have stronger compatibility and lower costs, suitable for a wider range of environments. Furthermore, these techniques can effectively reduce motion blur without changing the camera’s hardware configuration or shooting parameters. In addition, many deblurring algorithms can retain image details while processing blur, thus obtaining clear images under various lighting and contrast conditions. Therefore, deblurring technology is a very promising solution, especially suitable for applications such as home service robots that need to obtain clear images in various environmental conditions. Deblurring, also called image sharpening, is a typical operation in image processing, which includes differential methods and filtering methods. Traditional algorithms employ mathematical principles to sharpen blurred images according to their own characteristics.

However, these algorithms have the disadvantages of low generalization and adaptability. In contrast, image deblurring algorithms based on deep learning have excellent sharpening effects. For example, the hierarchical polyhedral network DMPHN [[8](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B8-sensors-23-09482)] calculated the standard performance indicators (PSNR and SSIM) on the GoPro dataset as 30.25 and 0.935, respectively. Compared with the former, the multiscale input and multiscale output MIMO-UNet+ based on a single U-Net [[9](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B9-sensors-23-09482)], the multiscale hierarchical network MSSNet [[10](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B10-sensors-23-09482)], and the multistage network HINet [[11](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B11-sensors-23-09482)] based on the Half-Instance Normalization module achieved PSNR improvements of 1.25, 1.95, and 1.51 and SSIM improvements of 0.022, 0.018, and 0.024, respectively. The above algorithms have all demonstrated good performance in their adaptability to complex images in multiple scenes. However, they fail to consider the motion information of the robot’s body, which requires a complex network structure to ensure adaptability. As a result, they rely on the support of a large server to ensure the real-time performance of the calculation (for example, the processing time of DMPHN on a single NVIDIA Tesla P100 GPU is 30 ms, the inference time of MIMO-UNet+ on an NVIDIA Titan XP is 17 ms, the inference time of HINet on an NVIDIA Tesla V100 GPU is 27 ms, and the processing time of MSSNet on an NVIDIA RTX 3090 GPU is as high as 104 ms). However, service robots usually use an embedded processing platform, which cannot supply the computing resource required by the above algorithms.

After deblurring, service robots need to detect the target object in an image. Traditional object detection algorithms usually scan the entire image using a multiscale sliding window; find the Region of Interest (ROI) [[12](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B12-sensors-23-09482)] that may exist in the image; use the Scale-Invariant Feature Transformation (SIFT) [[13](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B13-sensors-23-09482)], the Histogram of Oriented Gradient (HOG) [[14](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B14-sensors-23-09482)], the Deformable Part-based Model (DPM) [[15](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B15-sensors-23-09482)], the Local Binary Pattern (LBP) [[16](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B16-sensors-23-09482)], and other algorithms to extract the expected features of the ROI; and, finally, use Support Vector Machine (SVM) [[17](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B17-sensors-23-09482),[18](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B18-sensors-23-09482),[19](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B19-sensors-23-09482)], iterative adaptive boosting algorithm (Adaboost) [[20](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B20-sensors-23-09482)], or other classifiers to classify and recognize the object. Expected features are generally defined artificially, such as edge, color, shape, and size [[21](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B21-sensors-23-09482),[22](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B22-sensors-23-09482),[23](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B23-sensors-23-09482),[24](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B24-sensors-23-09482)]. The traditional object detection algorithms have pretty good accuracy in fixed scenes, but the accuracy depends on artificially designed features. Therefore, these algorithms were designed empirically and have poor generalization ability. They are suitable for specific scenarios and objects, but they cannot meet the complex scenarios and multitasking requirements of service robots in living environments.

In recent years, object-detection methods based on deep learning have been extensively studied and proved to be a promising solution. In 2014, an ROI-based convolutional network approach (R-CNN) was first introduced into the field of object detection [[25](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B25-sensors-23-09482)]. Subsequently, Fast R-CNN [[26](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B26-sensors-23-09482)], Faster R-CNN [[27](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B27-sensors-23-09482)], R-FCN [[28](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B28-sensors-23-09482)], SPP-Net [[29](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B29-sensors-23-09482)], VGGNet [[30](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B30-sensors-23-09482)], GoogLeNet [[31](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B31-sensors-23-09482)], and other two-stage model object detection methods had been proposed. In 2016, an end-to-end object detection algorithm (YOLO) was proposed [[32](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B32-sensors-23-09482)], and YOLOv2 [[33](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B33-sensors-23-09482)], YOLOv3 [[34](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B34-sensors-23-09482)], SSD [[35](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B35-sensors-23-09482)], and other algorithms subsequently appeared, all belonging to single-stage models. In order to detect small objects such as partially occluded bottles in unstructured scenes, various algorithms had been employed by researchers including R-CNN, Fast R-CNN, Faster R-CNN, YOLO, YOLOv2, and SSD and their precisions reached 36.8%, 38.3%, 52.1%, 22.7%, 51.8%, and 53.2%, respectively. Although the above methods realize object detection, their precisions are not high enough for applications. The main reason is that the problem of small and locally occluded target items has not yet been resolved.

Addressing the problem of small and partially occluded object detection, a dynamic discarding technique based on YOLOv3 was deployed using the NVIDIA Jetson Xavier platform for an autonomous vehicle. This algorithm could complete long-distance pedestrian detection, and the detection time of each image was 41.66 ms. This technique can only run on a Jetson GPU and does not utilize TensorRT [[36](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B36-sensors-23-09482)]. The mobile robot equipped with an NVIDIA Jetson TX2 embedded device used the YOLOv2 algorithm to detect a table (table foot) in the office setting, achieving a precision of 69% and an image detection time of 232 ms [[37](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B37-sensors-23-09482)]. An improved Tiny-YOLO algorithm using the ROBO structure and pruning technology was applied to the Nao robot for object detection, with a mAP of 78.75% and an inference time of 172 ms for each image [[38](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B38-sensors-23-09482)]. A picking robot embedded with the Adreno 640 platform used a Light-YOLOv3 algorithm with a multiscale context aggregation structure and a lightweight network to detect green mangoes with an F1% of 97.7 and a model inference efficiency of 62.5 ms [[39](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B39-sensors-23-09482)]. Another picking robot equipped with a Snapdragon 865-embedded device, running on an improved YOLOv3 algorithm was used to achieve green peach detection, with a precision of 97.3% and a test time of 60.1 ms pre-image [[40](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B40-sensors-23-09482)]. The accuracies of the above methods had been greatly improved, and the detection of small and partially occluded objects had also achieved good results. At the same time, the efficiency of model reasoning under limited computing resources had also been optimized to a certain extent. However, the inference time of these algorithms still cannot meet the processing speed requirement of 30 FPS for real-time vision. Moreover, the above object detection algorithms cannot simultaneously meet the real-time object detection requirements of service robots in complex indoor scenes.

Researchers have also combined object detection with image deblurring algorithms. A Deblur-YOLO algorithm for detecting objects from motion-blurred images executed on a single NVIDIA GTX 1080Ti GPU was developed, with a mAP of 47.5%, a model inference efficiency of 77.2 ms, and a model size of 12.9 MB [[41](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B41-sensors-23-09482)]. Motion-blurred images were also processed using the DeblurGANv2-InceptionResNet [[42](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B42-sensors-23-09482)], SRN [[43](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B43-sensors-23-09482)], DeepDeblur [[44](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B44-sensors-23-09482)], and DynamicDeblur [[45](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B45-sensors-23-09482)] algorithms, respectively. The time consumptions were 158 ms, 379 ms, 1550 ms, and 1525 ms, respectively, and the model sizes were 233.0 MB, 86.9 MB, 47.4 MB, and 47.8 MB, respectively. Subsequently, the YOLOv3 algorithm was used to perform object detection on clear images and motion-deblurred images by using the DeblurGANv2-InceptionResNet [[42](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B42-sensors-23-09482)], SRN [[43](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B43-sensors-23-09482)], DeepDeblur [[44](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B44-sensors-23-09482)], and DynamicDeblur [[45](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10708789/#B45-sensors-23-09482)] algorithms, with a mAP of 58.5%, 42.0%, 52.3%, 51.7%, and 56.0%, respectively. However, these algorithms focused on the motion deblurring performance of images and did not pay attention to the real-time performance of both deblurring and object detection. As a result, these algorithms have complex model network structures and high inference delays, making it difficult to achieve dynamic and real-time object detection for home service robots.

Therefore, this paper proposes a dynamic and real-time object detection algorithm for home service robots based on DA-Multi-DCGAN and AT-LI-YOLO, which is able to achieve deblurring on different motion-state images based on GAN through its embedded dynamic adjustment mechanism and multiscale fusion structure. It also performs real-time object detection of target objects via the YOLOv3 model with an attention module embedded in AT-Resblock and a lightweight network unit named Lightblock. While improving the detection precision of small objects and locally occluded objects, this method can well meet the real-time requirement of object detection under the condition of limited computing resources. The main innovations and contributions of this paper are as follows:

* A DA-Multi-DCGAN with a dynamic adjustment mechanism combining PSNR, SSIM, and sharpness and a multimodal multiscale fusion structure taking robot motion and the surrounding environment information into account is proposed, which can realize the deblurring of the acquired images under different motion states and enhance the efficiency.
* An AT-Resblock embedded with an attention module is designed, which can fully extract features such as color and texture contours of objects, highlight key areas, integrate salient features, and improve the sensitivity and precision of small and partially occluded objects. Based on a partial depthwise separable convolution that replaces the standard convolution to improve model detection speed, a lightweight network unit named Lightblock is developed, which reduces the number of network parameters and computational complexity, and improves computational efficiency.
* We analyzed the daily manipulating tasks of older people and children, summarized 17 kinds of frequently used daily necessities, and constructed a dataset named Grasp-17 for object detection of service robots.
* The TensorRT neural network inference engine was run on the home service robot prototype equipped with NVIDIA Jeston AGX Xavier, and the inference efficiency was improved by three times under the same model, which achieved the dynamic object detection time of 29 ms, demonstrating the real-time nature of the proposed method in service robots during motion.

.

# Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research

# Diabetic Retinopathy is the most prevalent cause of avoidable vision impairment, mainly affecting the working-age population in the world. Recent research has given a better understanding of the requirement in clinical eye care practice to identify better and cheaper ways of identification, management, diagnosis and treatment of retinal disease. The importance of diabetic retinopathy screening programs and difficulty in achieving reliable early diagnosis of diabetic retinopathy at a reasonable cost needs attention to develop computer-aided diagnosis tool. Computer-aided disease diagnosis in retinal image analysis could ease mass screening of populations with diabetes mellitus and help clinicians in utilizing their time more efficiently. The recent technological advances in computing power, communication systems, and machine learning techniques provide opportunities to the biomedical engineers and computer scientists to meet the requirements of clinical practice. Diverse and representative retinal image sets are essential for developing and testing digital screening programs and the automated algorithms at their core. To the best of our knowledge, IDRiD (Indian Diabetic Retinopathy Image Dataset), is the first database representative of an Indian population. It constitutes typical diabetic retinopathy lesions and normal retinal structures annotated at a pixel level. The dataset provides information on the disease severity of diabetic retinopathy, and diabetic macular edema for each image. This makes it perfect for development and evaluation of image analysis algorithms for early detection of diabetic retinopathy.

Diabetic Retinopathy (DR) is the result of microvascular retinal changes triggered by diabetes and it is the most common leading cause of preventable blindness in the working-age population in the world [[**1**](https://www.mdpi.com/2306-5729/3/3/25#B1-data-03-00025),[**2**](https://www.mdpi.com/2306-5729/3/3/25#B2-data-03-00025)]. Whereas, Diabetic Macular Edema (DME) is a complication associated with DR, characterized by accumulation of fluid or retinal thickening that can occur at any stage of DR [[**3**](https://www.mdpi.com/2306-5729/3/3/25#B3-data-03-00025),[**4**](https://www.mdpi.com/2306-5729/3/3/25#B4-data-03-00025)]. International Council of Ophthalmology (ICO) report [[**5**](https://www.mdpi.com/2306-5729/3/3/25#B5-data-03-00025)] indicate that 1 out of 3 individuals affected with diabetes had some form of DR and also highlighted that 1 in 10 had vision-threatening DR. In India it is the sixth common cause of blindness [[**6**](https://www.mdpi.com/2306-5729/3/3/25#B6-data-03-00025)].

DR is referred as a clinical diagnosis, depicted by the presence (see [**Figure 1**](https://www.mdpi.com/2306-5729/3/3/25#fig_body_display_data-03-00025-f001)) of one or more several retinal lesions like microaneurysms, hemorrhages, hard exudates and soft exudates [[**7**](https://www.mdpi.com/2306-5729/3/3/25#B7-data-03-00025)].

Early diagnosis and treatment of DR can prevent vision loss [[**8**](https://www.mdpi.com/2306-5729/3/3/25#B8-data-03-00025)]. Hence, diabetic patients are referred to do a regular biannual or annual follow-up and frequent consultation for the screening of their retina [[**9**](https://www.mdpi.com/2306-5729/3/3/25#B9-data-03-00025)]. The elimination of preventable visual impairment is mainly dependent on the pool of expert clinicians and basic health care infrastructure essential for the treatment of the eye [[**10**](https://www.mdpi.com/2306-5729/3/3/25#B10-data-03-00025),[**11**](https://www.mdpi.com/2306-5729/3/3/25#B11-data-03-00025)]. In the Indian subcontinent, against national eye care experts: population ratio of 1:107,000, in various regions this ratio is 1:9000 whereas in some other parts there is only one eye care expert for 608,000 population [[**12**](https://www.mdpi.com/2306-5729/3/3/25#B12-data-03-00025),[**13**](https://www.mdpi.com/2306-5729/3/3/25#B13-data-03-00025)]. Due to the large number of people that require a continuous follow-up and shortage of ophthalmologists, management of DR needs attention to develop computer-aided diagnosis tool [[**14**](https://www.mdpi.com/2306-5729/3/3/25#B14-data-03-00025),[**15**](https://www.mdpi.com/2306-5729/3/3/25#B15-data-03-00025)]. The recent technological advances in computing power, communication systems, and machine learning techniques provide opportunities to the biomedical engineers and computer scientists to meet the requirements of clinical practice [[**16**](https://www.mdpi.com/2306-5729/3/3/25#B16-data-03-00025),[**17**](https://www.mdpi.com/2306-5729/3/3/25#B17-data-03-00025)]. The raw images with ground truths facilitates the scientific community for development, validation, comparison and aid in the further improvement of DR lesion detection algorithms used in clinical application [[**18**](https://www.mdpi.com/2306-5729/3/3/25#B18-data-03-00025)]. Precise pixel level annotation of abnormalities associated with DR like microaneurysms, soft exudates, hard exudates and hemorrhages is invaluable resource for performance evaluation of individual lesion segmentation techniques. Whereas, the reliable information about disease severity level of DR, and DME is useful in development and evaluation of image analysis and retrieval algorithms for early detection of the disease [[**19**](https://www.mdpi.com/2306-5729/3/3/25#B19-data-03-00025)].

This dataset was available as a part of “Diabetic Retinopathy: Segmentation and Grading Challenge ([**http://biomedicalimaging.org/2018/challenges/**](http://biomedicalimaging.org/2018/challenges/))” organized in conjunction with IEEE International Symposium on Biomedical Imaging (ISBI-2018), Washington D.C. The data challenge was hosted on Grand Challenges in Biomedical Imaging Platform [[**20**](https://www.mdpi.com/2306-5729/3/3/25#B20-data-03-00025)]. Information about specifications and data accessibility is provided in the [**Table 1**](https://www.mdpi.com/2306-5729/3/3/25#table_body_display_data-03-00025-t001).

# Deep Convolution Neural Network sharing for the multi-label images classification

# Addressing issues related to multi-label classification is relevant in many fields of applications. In this work. We present a multi-label classification architecture based on Multi-Branch [Neural Network Model](https://www.sciencedirect.com/topics/computer-science/neural-network-model) (MBNN) that permits the network to encode data from multiple semi-parallel [subnetworks](https://www.sciencedirect.com/topics/computer-science/subnetwork) or layers outputs separately. Different types of [neural networks](https://www.sciencedirect.com/topics/computer-science/neural-network) can be used in the MBNN, but the proposal is made with [Convolutional Neural Networks](https://www.sciencedirect.com/topics/computer-science/convolutional-neural-network) subnetworks, trained, and joined in classifying the outputs (i.e., labels). The proposed work makes it possible to perform incremental changes on existing [Multitask Learning](https://www.sciencedirect.com/topics/computer-science/multitask-learning) architectures for an adaptation to the multi-label classification. These transformations lead us to define two new architectures (neural network multi-outputs and neural network multi-features) using the feature extractors from the pre-trained neural networks. The empirical and statistical results verify that the proposed multibranch neural network architecture performs better than other simple multi-label classification architectures. Later, the “network with multi-features” obtained the highest classification score than other [deep neural networks](https://www.sciencedirect.com/topics/computer-science/deep-neural-network) with 83.31% of the f1-score for the Amazon rainforest dataset. The f1-score values are 88.81% for Pascal VOC 2007 dataset, 87.71% for Nuswide, and 88.64% for Pascal VOC 2012.

[Deep learning](https://www.sciencedirect.com/topics/computer-science/deep-learning) is a subfield of [machine learning](https://www.sciencedirect.com/topics/computer-science/machine-learning). It has been widely used for [image classification](https://www.sciencedirect.com/topics/computer-science/image-classification) ([Abdullahi et al., 2017](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b1), [Seo and Shin, 2019](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b80), [Shiqi et al., 2017](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b84)) with good performances, in segmentation ([Garcia-Garcia et al., 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b39), [Längkvist et al., 2016](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b56)), object detection ([Cheng and Han, 2016](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b21), [Gopalakrishnan et al., 2017](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b42), remote sensing [Atharva et al., 2017](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b8), [Fotso Kamga, 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b37)), and in many other applications ([Shahin et al., 2017](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b82), [Wason, 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b95)).

In recent years, the problem of multi-label classification (MLC) has emerged in various application domains including image annotation ([Boutell, Luo, Shen, & Brown, 2004](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b17)); ([Changhu, Shuicheng, Lei, & Hong-Jiang, 2009](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b19)), text categorization, gene function annotation ([Li et al., 2010](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b59), [Yang et al., 2019](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b101)) and so on. Each instance to be classified is characterized by more than one class label (categories), proving this topic’s importance for many researchers. These applications use different techniques to solve multi-label classification problems: [binary classification](https://www.sciencedirect.com/topics/computer-science/binary-classification), information analysis, feature selection, and machine learning.

The multi-label classification aims to offer learning opportunities to produce label subsets as output ([He et al., 2019](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b44), [Huang et al., 2017](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b49), [Pereira et al., 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b74)). The simplest way to classify a multi-label dataset is to convert it into a single-label dataset with multiple classes. However, in some cases, it is necessary to consider the correlation between labels to improve the performance task ([Weng et al., 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b97), [Wosiak et al., 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b98)). Accordingly, there are two significant groups of multi-label classification methods. Accordingly, there are two major categories of multi-label classification methods ([Tsoumakas & Katakis, 2007](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b91)): (i) problem transformation and (ii) [algorithm adaptation](https://www.sciencedirect.com/topics/computer-science/adaptation-algorithm) methods.

The most popular method is problem transformation, where a multi-label dataset is converted to a single-label dataset using simple operations such as copy, copy-weight, select-max, select-min, select-random, and ignore. For example, Binary Relevance (BR) is a popular method of problem transformation that involves learning several [binary classifiers](https://www.sciencedirect.com/topics/computer-science/binary-classifier), i.e., one classifier for each distinctive label in the finite set of labels.

The algorithm adaptation methods support the idea of an extension for specific learning algorithms (e.g., a probabilistic [generative model](https://www.sciencedirect.com/topics/computer-science/generative-model), [conditional random fields](https://www.sciencedirect.com/topics/computer-science/conditional-random-field), a deconvolution approach, etc.) to manipulate multi-label data to conduct multi-label predictions directly. An example of algorithm adaptation methods is conditional random fields ([Ghamrawi & McCallum, 2005](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b40)), which use two [graphical models](https://www.sciencedirect.com/topics/computer-science/graphical-model) for setting label co-occurrences. The first one concerns collective multi-labels with co-occurrence patterns between labels. In contrast, the second one concerns collective multi-labels with features and seeks to capture an individual feature’s influence on the co-occurrence probabilities of a pair of labels.

In this work, we propose a new approach to multi-label classification by designing a Multi-Branch [Neural Network](https://www.sciencedirect.com/topics/computer-science/neural-network) (MBNN). We try to address the following issues:

(a) how to transform a multi-label classification problem into a binary classification problem? (b) how, to solve each problem to find the relevant label classes?

As a solution to this problem, we use the divide-and-conquer technique, where each problem is defined by (i) building a multi-branch model (ii) each branch deals with a part of the problem, and (iii) we combine the result of each branch with building the final label space which is the result of the multi-label prediction.

[Multitask learning](https://www.sciencedirect.com/topics/computer-science/multitask-learning) aims at solving more than one problem (or task) simultaneously. Therefore, its learning process translates difficulties into multiple tasks. Initially, parameters are learned by training on source tasks (feature extractors). In the next step, the parameters obtained are used to learn the target task using its associated datasets. This strengthens the network’s ability against overfitting compared to training the model on each task separately. Recently, deep multitask learning is used to solve various [natural language processing](https://www.sciencedirect.com/topics/computer-science/natural-language-processing) problems, sentiment and [emotion analysis](https://www.sciencedirect.com/topics/computer-science/emotion-analysis) ([Akhtar et al., 2020](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b4), [Akhtar et al., 2019](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b5)), in computer vision ([Yidong et al., 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b103), [Zhuang et al., 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b112)), or multimodal learning ([Baltrusaitis, Ahuja, & Morency, 2019](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b9)).

Shared representations improve data efficiency. So, it can potentially increase the speed of learning for related or downstream tasks, thereby mitigating the weaknesses of [deep neural networks](https://www.sciencedirect.com/topics/computer-science/deep-neural-network), such as their need for large-scale training data and high computational power. These effects, however, have not been easy to achieve and require [architectural design](https://www.sciencedirect.com/topics/computer-science/architectural-design) ([Liu et al.](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b62), [Ruder, 2017](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b79), [Zhao et al., 2018](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b110)) and optimization ([Gao et al., 2020](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b38), [Karasu et al., 2020](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b54), [Liu et al., 2019](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b63), [Yu et al., 2020a](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b106)). It is therefore challenging to avoid inter-task conflicts and learn about the relationships between tasks ([Song et al., 2019](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b86), [Strezoski et al., 2019](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b88)). In addition, optimization methods and [image processing](https://www.sciencedirect.com/topics/computer-science/image-processing) significantly impact the performance of [deep learning models](https://www.sciencedirect.com/topics/computer-science/deep-learning-model) ([Sezer & Altan, 2021](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b81)). Therefore, the combination of models can also be relevant, such as a hybrid model comprising two-dimensional (2D) curvelet transformation, chaotic salp swarm algorithms (CSSA), and [deep learning techniques](https://www.sciencedirect.com/topics/computer-science/deep-learning-technique) ([Altan & Karasu, 2020](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b6)). In the same vein, engaging multi-objective optimization techniques may be helpful for wrapper-based feature selection in forecasting models for challenging tasks with nonlinear dynamics ([Karasu et al., 2020](https://www.sciencedirect.com/science/article/pii/S2666827022000974#b54)).

Despite progress, optimization techniques or learning relationships hardly converge in the neural network if it is too deep with billions of parameters to learn. These reasons have led us to consider a training process based on the pretrained network’s architecture. These reasons have led us to consider a training process based on the pretrained network’s architecture. These reasons have led us to consider a training process based on the pretrained network’s architecture. The main idea is to show the possibility of integrating [multiple classifiers](https://www.sciencedirect.com/topics/computer-science/multiple-classifier) or feature extractors to obtain a strong classifier with better generalization than each component. Moreover, in comparisons of classifiers, the accuracy and efficiency of the experimental results are proven in statistical tests, something less frequent in work.

On the one hand, the empirical results show that the proposed architecture’s performance is better than that of the simple multi-label [classification model](https://www.sciencedirect.com/topics/computer-science/classification-models). However, on the other hand, if our architectures are executed simultaneously on a single [computational node](https://www.sciencedirect.com/topics/computer-science/computational-node), it may cause latency when the number of feature extractors increases or several binary classifiers are defined. Finally, the main contributions of this study are:

* We recommend a multitask learning framework called multibranch neural network for multi-label classification. We redefined two new architectures (neural network multi-outputs and neural network multi-features) of the multibranch neural network using the feature extractors from the pre-trained neural networks.
* Furthermore, we recommend using pre-trained neural networks that will be translated as an implicit [data augmentation](https://www.sciencedirect.com/topics/computer-science/data-augmentation).
* The performance of the proposed models is demonstrated on 4 data sets by empirical and statistical results.

# Unsupervised Domain Adaptation with Coupled Generative Adversarial Autoencoders

When large-scale annotated data are not available for certain image classification tasks, training a deep convolutional neural network model becomes challenging. Some recent domain adaptation methods try to solve this problem using generative adversarial networks and have achieved promising results. However, these methods are based on a shared latent space assumption and they do not consider the situation when shared high level representations in different domains do not exist or are not ideal as they assumed. To overcome this limitation, we propose a neural network structure called coupled generative adversarial autoencoders (CGAA) that allows a pair of generators to learn the high-level differences between two domains by sharing only part of the high-level layers. Additionally, by introducing a class consistent loss calculated by a stand-alone classifier into the generator optimization, our model is able to generate class invariant style-transferred images suitable for classification tasks in domain adaptation. We apply CGAA to several domain transferred image classification scenarios including several benchmark datasets. Experiment results have shown that our method can achieve state-of-the-art classification results.

Large-scale well-annotated datasets such as Microsoft COCO [[**1**](https://www.mdpi.com/2076-3417/8/12/2529#B1-applsci-08-02529)], ImageNet [[**2**](https://www.mdpi.com/2076-3417/8/12/2529#B2-applsci-08-02529)] and KITTI [[**3**](https://www.mdpi.com/2076-3417/8/12/2529#B3-applsci-08-02529)] have played a vital role in the recent success of deep learning based models on computer vision tasks such as image classification, target detection, semantic segmentation and so on. However, models trained with large datasets still cannot generalize well to novel datasets when these datasets have different feature distributions. The typical solution is to further fine-tune these models on the task specific datasets. However, creating such datasets can be expensive and time-consuming. Unsupervised domain adaptation offers a solution to this problem by learning a mapping between a labeled dataset (source domain) and an unlabeled dataset (target domain) or by learning domain invariant features. Conventional domain adaptation approaches for image classification are usually developed in two separate steps: designing and extracting fixed features and then training models to reduce their differences in either the marginal distributions or the conditional distributions between domains [[**4**](https://www.mdpi.com/2076-3417/8/12/2529#B4-applsci-08-02529),[**5**](https://www.mdpi.com/2076-3417/8/12/2529#B5-applsci-08-02529),[**6**](https://www.mdpi.com/2076-3417/8/12/2529#B6-applsci-08-02529),[**7**](https://www.mdpi.com/2076-3417/8/12/2529#B7-applsci-08-02529)]. Recent deep learning based domain adaptation approaches avoid the difficulty of feature design by extracting features automatically through convolutional neural networks [[**8**](https://www.mdpi.com/2076-3417/8/12/2529#B8-applsci-08-02529),[**9**](https://www.mdpi.com/2076-3417/8/12/2529#B9-applsci-08-02529),[**10**](https://www.mdpi.com/2076-3417/8/12/2529#B10-applsci-08-02529),[**11**](https://www.mdpi.com/2076-3417/8/12/2529#B11-applsci-08-02529),[**12**](https://www.mdpi.com/2076-3417/8/12/2529#B12-applsci-08-02529),[**13**](https://www.mdpi.com/2076-3417/8/12/2529#B13-applsci-08-02529)].

Among all kinds of deep neural network based domain adaptation approaches, generative adversarial network (GAN) [[**14**](https://www.mdpi.com/2076-3417/8/12/2529#B14-applsci-08-02529)] has become a popular branch. A typical GAN trains a generator and a discriminator to compete against each other. The generator is trained to produce synthetic images as real as possible, whereas the discriminator is trained to distinguish the synthetic and real images. When applying GAN to domain adaptation for image classification, there are two major types of approaches. The first type trains a GAN to generate unlabeled target domain images, thus enlarging the data volume to train a more robust image classifier [[**15**](https://www.mdpi.com/2076-3417/8/12/2529#B15-applsci-08-02529),[**16**](https://www.mdpi.com/2076-3417/8/12/2529#B16-applsci-08-02529),[**17**](https://www.mdpi.com/2076-3417/8/12/2529#B17-applsci-08-02529)]. In these methods, the training strategy of the final classifier need to be carefully designed since the newly generated images have no label. The other type of approaches generate labeled target domain images directly by transferring the source domain images into target domain style and have achieved some state-of-the-art results, such as CoGAN [[**18**](https://www.mdpi.com/2076-3417/8/12/2529#B18-applsci-08-02529)] and UNIT [[**19**](https://www.mdpi.com/2076-3417/8/12/2529#B19-applsci-08-02529)]. These methods are based on the shared latent space assumption, which assumes that the differences of the source domain and the target domain are primarily low-level, and that the two domains share a common high-level latent space. This assumption works well for simple scenarios such as digits adaptation between MNIST [[**20**](https://www.mdpi.com/2076-3417/8/12/2529#B20-applsci-08-02529)] and USPS [[**21**](https://www.mdpi.com/2076-3417/8/12/2529#B21-applsci-08-02529)] but faces challenges when the semantic features are more complex. When shared high-level latent space in different domains does not exist or such latent space is not as ideal as assumed, these methods will fail [[**18**](https://www.mdpi.com/2076-3417/8/12/2529#B18-applsci-08-02529)].

In this paper, we propose an unsupervised domain adaptation method for image classification by combining generative adversarial networks with autoencoders. We call our proposed network architecture Coupled Generative Adversarial Autoencoders (CGAA). Our work is perhaps most similar to CoGAN and UNIT, but we try to solve the aforementioned shortcomings of these methods by the following designs: CGAA consists of a pair of generative adversarial networks (GAN) and a domain adaptive classifier. The architecture of the generator in GAN is designed based on the autoencoder. During training, part of the layers in the generators are forced to share their weights, which gives our model the ability to learn the domain transformation in an unsupervised manner and generate synthetic target domain images with label. By decoupling the highest level layer, we give our model the capacity to tolerant the differences of high-level features between the domains. The classifier provides a class-invariant loss to help the generator produce more suitable images for the classification task in domain adaptation. The main contributions of this work are:

* We propose an unsupervised domain adaptation method for image classification. Our method trains a pair of coupled generative adversarial networks in which the generator has an encoder-decoder structure.
* We force part of the layers in the generator to share weights during training to generate labeled synthetic images, and make the highest level layer decoupled for different high-level representations.
* We introduce a class consistent loss into the GAN training, which is calculated from the output of a stand-alone domain adaptive classifier. It can help the generator to generate more suitable images for domain adaptation.

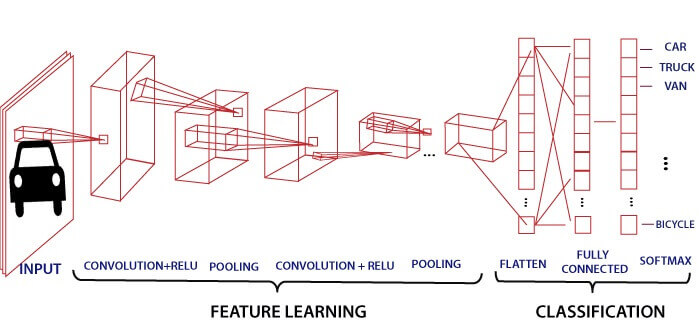
**IMPLEMENTATION**

# CONVOLUTIONAL NEURAL NETWORK

**Convolutional Neural Network** is one of the main categories to do image classification and image recognition in neural networks. Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolutional neural networks are widely used.

CNN takes an image as input, which is classified and process under a certain category such as dog, cat, lion, tiger, etc. The computer sees an image as an array of pixels and depends on the resolution of the image. Based on image resolution, it will see as **h \* w \* d**, where h= height w= width and d= dimension. For example, An RGB image is **6 \* 6 \* 3** array of the matrix, and the grayscale image is **4 \* 4 \* 1** array of the matrix.

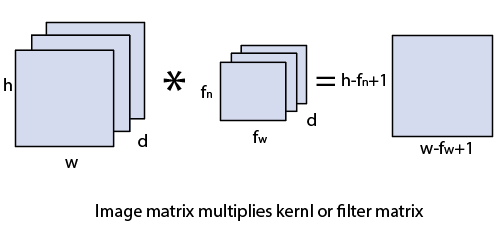
In CNN, each input image will pass through a sequence of convolution layers along with pooling, fully connected layers, filters (Also known as kernels). After that, we will apply the Soft-max function to classify an object with probabilistic values 0 and 1.



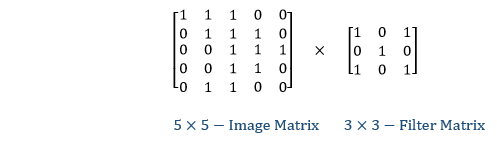
## Convolution Layer

Convolution layer is the first layer to extract features from an input image. By learning image features using a small square of input data, the convolutional layer preserves the relationship between pixels. It is a mathematical operation which takes two inputs such as image matrix and a kernel or filter.

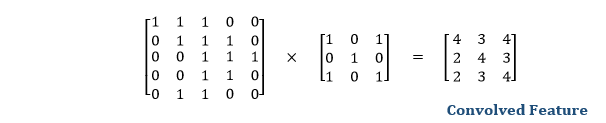
* The dimension of the image matrix is **h×w×d**.
* The dimension of the filter is **fh×fw×d**.
* The dimension of the output is **(h-fh+1)×(w-fw+1)×1**.



Let's start with consideration a 5\*5 image whose pixel values are 0, 1, and filter matrix 3\*3 as:



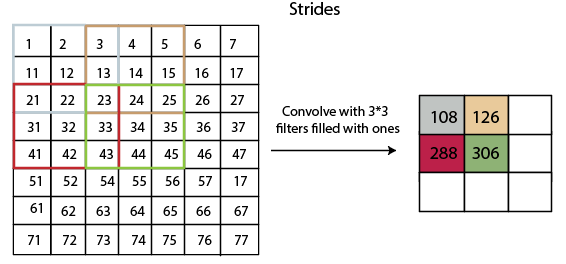
The convolution of 5\*5 image matrix multiplies with 3\*3 filter matrix is called "**Features Map**" and show as an output.



Convolution of an image with different filters can perform an operation such as blur, sharpen, and edge detection by applying filters.

## Strides

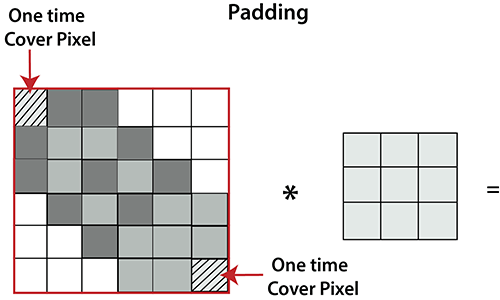
Stride is the number of pixels which are shift over the input matrix. When the stride is equaled to 1, then we move the filters to 1 pixel at a time and similarly, if the stride is equaled to 2, then we move the filters to 2 pixels at a time. The following figure shows that the convolution would work with a stride of 2.



## Padding

Padding plays a crucial role in building the convolutional neural network. If the image will get shrink and if we will take a neural network with 100's of layers on it, it will give us a small image after filtered in the end.

If we take a three by three filter on top of a grayscale image and do the convolving then what will happen?



It is clear from the above picture that the pixel in the corner will only get covers one time, but the middle pixel will get covered more than once. It means that we have more information on that middle pixel, so there are two downsides:

* Shrinking outputs
* Losing information on the corner of the image.

To overcome this, we have introduced padding to an image. **"Padding is an additional layer which can add to the border of an image."**

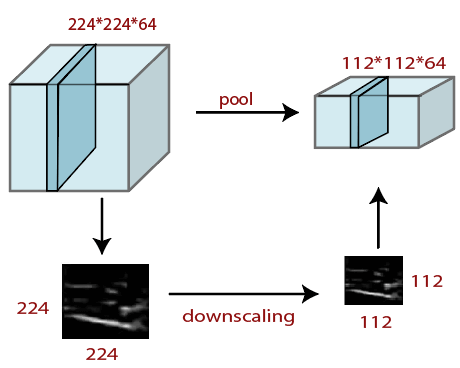
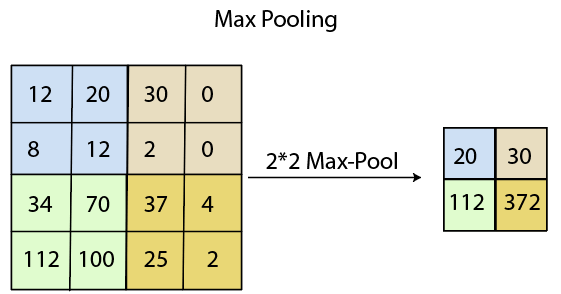
## Pooling Layer

Pooling layer plays an important role in pre-processing of an image. Pooling layer reduces the number of parameters when the images are too large. Pooling is "**downscaling**" of the image obtained from the previous layers. It can be compared to shrinking an image to reduce its pixel density. Spatial pooling is also called downsampling or subsampling, which reduces the dimensionality of each map but retains the important information. There are the following types of spatial pooling:

### Max Pooling

Max pooling is a **sample-based discretization process**. Its main objective is to downscale an input representation, reducing its dimensionality and allowing for the assumption to be made about features contained in the sub-region binned.

Max pooling is done by applying a max filter to non-overlapping sub-regions of the initial representation.



### Average Pooling

Down-scaling will perform through average pooling by dividing the input into rectangular pooling regions and computing the average values of each region.

**Syntax**

layer = averagePooling2dLayer(poolSize)

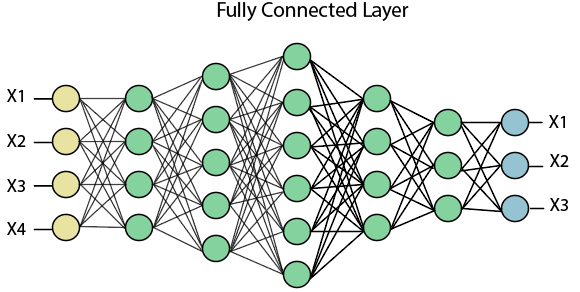
layer = averagePooling2dLayer(poolSize,Name,Value)

### Sum Pooling

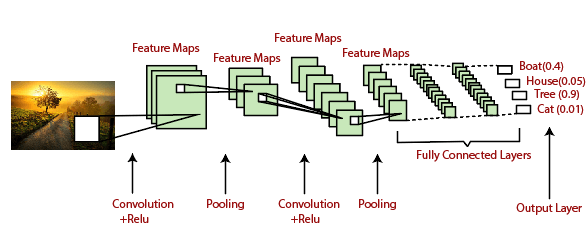
The sub-region for **sum pooling** or **mean pooling** are set exactly the same as for **max-pooling** but instead of using the max function we use sum or mean.

## Fully Connected Layer

The fully connected layer is a layer in which the input from the other layers will be flattened into a vector and sent. It will transform the output into the desired number of classes by the network.



In the above diagram, the feature map matrix will be converted into the vector such as **x1, x2, x3... xn** with the help of fully connected layers. We will combine features to create a model and apply the activation function such as **softmax** or **sigmoid** to classify the outputs as a car, dog, truck, etc.



**LOCAL BINARY PATTERN**

Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis. Perhaps the most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes caused, for example, by illumination variations. Another important property is its computational simplicity, which makes it possible to analyze images in challenging real-time settings.

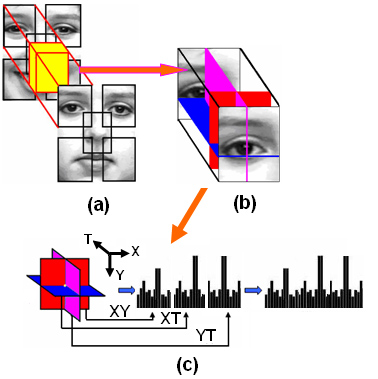


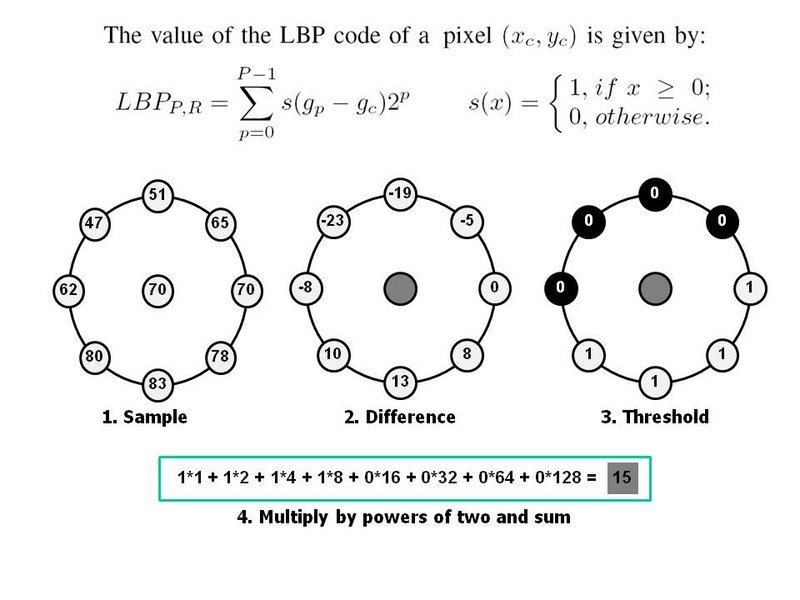
Fig : Description of facial expressions with local binary patterns.

**LBP IN SPATIAL DOMAIN**

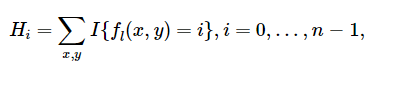
The basic idea for developing the LBP operator was that two-dimensional surface textures can be described by two complementary measures: local spatial patterns and gray scale contrast. The original LBP operator (Ojala et al. 1996) forms labels for the image pixels by thresholding the 3 x 3 neighborhood of each pixel with the center value and considering the result as a binary number. The histogram of these 28 = 256 different labels can then be used as a texture descriptor. This operator used jointly with a simple local contrast measure provided very good performance in unsupervised texture segmentation (Ojala and Pietikäinen 1999). After this, many related approaches have been developed for texture and color texture segmentation.

The LBP operator was extended to use neighborhoods of different sizes (Ojala et al. 2002). Using a circular neighborhood and bilinearly interpolating values at non-integer pixel coordinates allow any radius and number of pixels in the neighborhood. The gray scale variance of the local neighborhood can be used as the complementary contrast measure. In the following, the notation (P,R) will be used for pixel neighborhoods which means P sampling points on a circle of radius of R. See Fig. 2 for an example of LBP computation. Another extension to the original operator is the definition of so-called uniform patterns, which can be used to reduce the length of the feature vector and implement a simple rotation-invariant descriptor. This extension was inspired by the fact that some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform whereas the patterns 11001001 (4 transitions) and 01010010 (6 transitions) are not. In the computation of the LBP labels, uniform patterns are used so that there is a separate label for each uniform pattern and all the non-uniform patterns are labeled with a single label. For example, when using (8,R) neighborhood, there are a total of 256 patterns, 58 of which are uniform, which yields in 59 different labels.

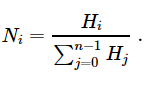
Ojala et al. (2002) noticed in their experiments with texture images that uniform patterns account for a little less than 90% of all patterns when using the (8,1) neighborhood and for around 70% in the (16,2) neighborhood. Each bin (LBP code) can be regarded as a micro-texton. Local primitives which are codified by these bins include different types of curved edges, spots, flat areas etc. The following notation is used for the LBP operator: LBPP,Ru2. The subscript represents using the operator in a (P,R) neighborhood. Superscript u2 stands for using only uniform patterns and labeling all remaining patterns with a single label. After the LBP labeled image fl(x,y) has been obtained, the LBP histogram can be defined as



**Fig :** An example of LBP computation.

****

in which n is the number of different labels produced by the LBP operator, and I{A} is 1 if A is true and 0 if A is false. When the image patches whose histograms are to be compared have different sizes, the histograms must be normalized to get a coherent description:



**SPATIO TEMPORAL LBP**

The original LBP operator was defined to only deal with the spatial information. Later, it was extended to a spatiotemporal representation for dynamic texture analysis. For this purpose, the so-called Volume Local Binary Pattern (VLBP) operator was proposed (Zhao and Pietikäinen 2007). The idea behind VLBP consists of looking at dynamic texture as a set of volumes in the (X,Y,T) space where X and Y denote the spatial coordinates and T denotes the frame index (time). The neighborhood of each pixel is thus defined in three dimensional space. Then, similarly to LBP in spatial domain, volume textons can be defined and extracted into histograms. Therefore, VLBP combines motion and appearance together to describe dynamic textures.

To make VLBP computationally simple and easy to extend, an operator based on co-occurrences of local binary patterns on three orthogonal planes (LBP-TOP) was also introduced. LBP-TOP considers three orthogonal planes: XY, XT and YT, and concatenates local binary pattern co-occurrence statistics in these three directions as shown in Figure [1](http://www.scholarpedia.org/article/Local_Binary_Patterns#fig:Your_article_title_Main_Facial.jpg). The circular neighborhoods are generalized to elliptical sampling to fit to the space-time statistics.

Fig. 3 shows example images from three planes. The XY plane represents appearance information, while the XT plane gives a visual impression of one row changing in time and YT describes the motion of one column in temporal space. The LBP codes are extracted for all pixels from the XY, XT and YT planes, denoted as XY-LBP, XT-LBP and YT-LBP, and histograms from these planes are computed and concatenated into a single histogram. In such a representation, a dynamic texture is encoded by an appearance (XY-LBP) and two spatial temporal (XT-LBP and YT-LBP) co-occurrence statistics. Setting the radius in the time axis to be equal to the radius in the space axis is not reasonable for dynamic textures. So we have different radius parameters in space and time to set. In the XT and YT planes, different radii can be assigned to sample neighboring points in space and time. More generally, the radii in axes X, Y and T, and the number of neighboring points in the XY, XT and YT planes can also be different

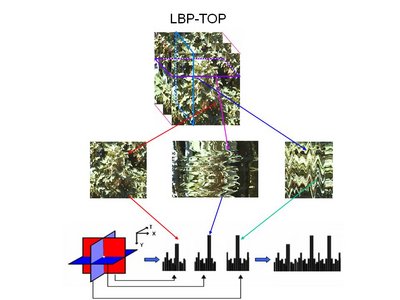


Fig: LBP from three orthogonal planes.

**EXTENSIONS AND APPLICATIONS**

The LBP methodology has led to significant progress in texture analysis. It is widely used all over the world both in research and applications. Due to its discriminative power and computational simplicity, the method has been very successful in many such computer [vision](http://www.scholarpedia.org/article/Vision) problems which were not earlier even regarded as texture problems, such as face analysis and motion analysis (Pietikäinen et al. 2011).

To increase the applicability of LBP, various extensions and modifications of it have been proposed. For example, Liao et al. (2009) proposed dominant local binary patterns which make use of the most frequently occurred patterns of LBP to improve the recognition accuracy. The use of interest region descriptors (such as [SIFT](http://www.scholarpedia.org/article/Scale_Invariant_Feature_Transform)) to various computer vision problems has been of great interest recently. For this purpose, a novel descriptor combining the strengths of SIFT and LBP was proposed (Heikkilä et al. 2009), in which center-symmetric local binary patterns (CS-LBP) were used to replace the gradient operator used by the SIFT operator. Mäenpää and Pietikäinen (2004) proposed an opponent color LBP, and investigated joint and separate use of color and texture in classification. The combination of the LBPs and Gabor features has been investigated (Tan et al. 2007, Wang et al. 2009).

The first texture-based method for background subtraction was proposed by Heikkilä and Pietikäinen (2006). Each pixel is modeled as a group of adaptive local binary pattern histograms that are calculated over a circular region around the pixel. The was shown to be tolerant to illumination variations, the multimodality of the background, and the introduction or removal of background objects. Furthermore, the method is capable for real-time processing. A preprocessing [algorithm](http://www.scholarpedia.org/article/Algorithm) based on the LBPs has been developed to handle variations in illumination in a face authentication system (Heusch et al. 2006). The use of LBPs in the recognition of actions was considered by (Kellokumpu et al. 2010). The use of LBPs in the facial age classification has been investigated (Wang et al. 2009). Other related LBP-based approaches to these problems have been proposed recently.

In addition to face and facial expression recognition, the LBP has also been used in many other applications of biometrics, including eye localization, iris recognition, fingerprint recognition, palmprint recognition, gait recognition and facial age classification.

**SYSTEM ARCHITECTURE**

**Training Process**

Retinal Database Images

Pre-process

(Noise Removal & Enhancement)

Training using CNN

Generation of Trained Model

**Testing Process**

Classified Results

(Different Grades of DR (0 to 4)

Input Retina Image

Preprocessing (Filtering by Median, Enhancement by CLAHE & Resize)

Classification using CNN Classifier

Trained Network Model

Performance Measure

**MODULES**

* Image Acquisition
* Pre-processing
* DR Grading
* Performance Measure

**MODULES DESCSRIPTION**

**Image Acquisition**

The Image Acquisition module serves as the initial step in our project, focusing on the retrieval of retinal images from the IDRiD database, which acts as the primary data source for our diabetic retinopathy grading system. Image acquisition is a crucial phase, as it provides the raw data necessary for further analysis and processing.

**Pre-processing**

The Preprocessing module plays a pivotal role in enhancing the quality and suitability of the retinal images for subsequent analysis. This module encompasses several key operations, including noise removal, contrast enhancement, and image resizing. It begins with the application of a median filter to effectively eliminate noise from the retinal images. Subsequently, the contrast of the images is improved using the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique. Finally, the images are resized to ensure uniformity and compatibility with the deep learning model.

**DR Grading**

* The Diabetic Retinopathy Grading module represents the core of our project, where the retinal images undergo grading to classify them into different diabetic retinopathy grades. This module leverages a Convolutional Neural Network (CNN) model, which has been meticulously trained using hyperparameters such as epochs, learning rate, dropout rate, and optimizer (SGDM).
* The grading process involves the categorization of retinal images into distinct grades, including 'no apparent retinopathy' (grade 0), 'mild NPDR' (grade 1), 'moderate NPDR' (grade 2), 'severe NPDR' (grade 3), and 'Proliferative Diabetic Retinopathy' (grade 4). The CNN model's ability to discern intricate patterns and features within the retinal images ensures precise and consistent grading, eliminating subjectivity and human bias.

**Performance Measure**

* Finally, our proposed model's performance was measured in terms of accuracy, error, precision, recall, specificity f1\_score and MCC performance metrics.
* Accuracy: It measures the analysis of TP and TN to the total no. of test images.

(1)

* Error Rate — what percentage of our prediction are wrong.

(2)

* Precision: It is the estimation analysis of true positive to the aggregate value of true positive and false positive rate. It is given in eqn. (3)

(3)

* Recall (Sensitivity): It is the estimation analysis of true positive rate to the aggregate value of the true positive and false negative rate. It is given in eqn. (4).

(4)

* Specificity - the proportion of categorized classes with negative class labels.

(5)

* F-Score: F-Measure is the harmonic mean of recall and precision. Precision and recall are given equal weight in the standard F-measure (F1).
* Matthews Correlation Coefficient (MCC): The MCC is in essence a correlation coefficient between the observed and predicted classifications.

(7)

The Performance Measure module evaluates the effectiveness and reliability of our proposed diabetic retinopathy grading system. It employs a comprehensive set of performance metrics to assess the system's performance:

Accuracy: Measures the overall correctness of the system's predictions.

Error Rate: Quantifies the percentage of incorrect predictions made by the system.

Precision: Evaluates the system's ability to make positive predictions correctly.

Recall (Sensitivity): Assesses the system's capacity to identify true positive cases.

Specificity: Measures the proportion of accurately identified negative cases.

F-Score: Computes the harmonic mean of precision and recall, providing a balanced performance measure.

Matthews Correlation Coefficient (MCC): Quantifies the correlation between observed and predicted classifications, accounting for both true and false positives and negatives.

These performance metrics collectively offer a comprehensive evaluation of the proposed system's accuracy, robustness, and reliability in diabetic retinopathy grading. They serve as essential indicators of the system's practical utility in clinical settings and its potential for improving patient care.

**SOFTWARE ENVIRONMENT**

**PYTHON**

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. An [interpreted language](https://en.wikipedia.org/wiki/Interpreted_language), Python has a design philosophy that emphasizes code [readability](https://en.wikipedia.org/wiki/Readability) (notably using [whitespace](https://en.wikipedia.org/wiki/Whitespace_character) indentation to delimit [code blocks](https://en.wikipedia.org/wiki/Code_block) rather than curly brackets or keywords), and a syntax that allows programmers to express concepts in fewer [lines of code](https://en.wikipedia.org/wiki/Source_lines_of_code) than might be used in languages such as [C++](https://en.wikipedia.org/wiki/C%2B%2B)or [Java](https://en.wikipedia.org/wiki/Java_(programming_language)). It provides constructs that enable clear programming on both small and large scales. Python interpreters are available for many [operating systems](https://en.wikipedia.org/wiki/Operating_system). [CPython](https://en.wikipedia.org/wiki/CPython), the [reference implementation](https://en.wikipedia.org/wiki/Reference_implementation) of Python, is [open source](https://en.wikipedia.org/wiki/Open_source) software and has a community-based development model, as do nearly all of its variant implementations. CPython is managed by the non-profit [Python Software Foundation](https://en.wikipedia.org/wiki/Python_Software_Foundation). Python features a [dynamic type](https://en.wikipedia.org/wiki/Dynamic_type) system and automatic [memory management](https://en.wikipedia.org/wiki/Memory_management). It supports multiple [programming paradigms](https://en.wikipedia.org/wiki/Programming_paradigm), including [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming), [imperative](https://en.wikipedia.org/wiki/Imperative_programming), [functional](https://en.wikipedia.org/wiki/Functional_programming) and [procedural](https://en.wikipedia.org/wiki/Procedural_programming), and has a large and comprehensive [standard library](https://en.wikipedia.org/wiki/Standard_library).

**Interactive Mode Programming**

Invoking the interpreter without passing a script file as a parameter brings up the following prompt −

$ python

Python 2.4.3 (#1, Nov 11 2010, 13:34:43)

[GCC 4.1.2 20080704 (Red Hat 4.1.2-48)] on linux2

Type "help", "copyright", "credits" or "license" for more information.

>>>

Type the following text at the Python prompt and press the Enter −

>>> print "Hello, Python!"

If you are running new version of Python, then you would need to use print statement with parenthesis as in print ("Hello, Python!");. However in Python version 2.4.3, this produces the following result −

Hello, Python!

**Script Mode Programming**

Invoking the interpreter with a script parameter begins execution of the script and continues until the script is finished. When the script is finished, the interpreter is no longer active.

Let us write a simple Python program in a script. Python files have extension .py. Type the following source code in a test.py file −

Live Demo

print "Hello, Python!"

We assume that you have Python interpreter set in PATH variable. Now, try to run this program as follows −

$ python test.py

This produces the following result −

Hello, Python!

Let us try another way to execute a Python script. Here is the modified test.py file −

Live Demo

#!/usr/bin/python

print "Hello, Python!"

We assume that you have Python interpreter available in /usr/bin directory. Now, try to run this program as follows −

$ chmod +x test.py # This is to make file executable

$./test.py

This produces the following result −

Hello, Python!

**Python Identifiers**

A Python identifier is a name used to identify a variable, function, class, module or other object. An identifier starts with a letter A to Z or a to z or an underscore (\_) followed by zero or more letters, underscores and digits (0 to 9).

Python does not allow punctuation characters such as @, $, and % within identifiers. Python is a case sensitive programming language. Thus, Manpower and manpower are two different identifiers in Python.

Here are naming conventions for Python identifiers −

Class names start with an uppercase letter. All other identifiers start with a lowercase letter.

Starting an identifier with a single leading underscore indicates that the identifier is private.

Starting an identifier with two leading underscores indicates a strongly private identifier.

If the identifier also ends with two trailing underscores, the identifier is a language-defined special name.

**Reserved Words**

The following list shows the Python keywords. These are reserved words and you cannot use them as constant or variable or any other identifier names. All the Python keywords contain lowercase letters only.

and exec not

assert finally or

break for pass

class from print

continue global raise

def if return

del import try

elif in while

else is with

except lambda yield

**Lines and Indentation**

Python provides no braces to indicate blocks of code for class and function definitions or flow control. Blocks of code are denoted by line indentation, which is rigidly enforced.

The number of spaces in the indentation is variable, but all statements within the block must be indented the same amount. For example −

if True:

print "True"

else:

print "False"

However, the following block generates an error −

if True:

print "Answer"

print "True"

else:

print "Answer"

print "False"

Thus, in Python all the continuous lines indented with same number of spaces would form a block. The following example has various statement blocks −

Note − Do not try to understand the logic at this point of time. Just make sure you understood various blocks even if they are without braces.

#!/usr/bin/python

import sys

try:

# open file stream

file = open(file\_name, "w")

except IOError:

print "There was an error writing to", file\_name

sys.exit()

print "Enter '", file\_finish,

print "' When finished"

while file\_text != file\_finish:

file\_text = raw\_input("Enter text: ")

if file\_text == file\_finish:

# close the file

file.close

break

file.write(file\_text)

file.write("\n")

file.close()

file\_name = raw\_input("Enter filename: ")

if len(file\_name) == 0:

print "Next time please enter something"

sys.exit()

try:

file = open(file\_name, "r")

except IOError:

print "There was an error reading file"

sys.exit()

file\_text = file.read()

file.close()

print file\_text

Multi-Line Statements

Statements in Python typically end with a new line. Python does, however, allow the use of the line continuation character (\) to denote that the line should continue. For example −

total = item\_one + \

item\_two + \

item\_three

Statements contained within the [], {}, or () brackets do not need to use the line continuation character. For example −

days = ['Monday', 'Tuesday', 'Wednesday',

'Thursday', 'Friday']

Quotation in Python

Python accepts single ('), double (") and triple (''' or """) quotes to denote string literals, as long as the same type of quote starts and ends the string.

The triple quotes are used to span the string across multiple lines. For example, all the following are legal −

word = 'word'

sentence = "This is a sentence."

paragraph = """This is a paragraph. It is

made up of multiple lines and sentences."""

Comments in Python

A hash sign (#) that is not inside a string literal begins a comment. All characters after the # and up to the end of the physical line are part of the comment and the Python interpreter ignores them.

Live Demo

#!/usr/bin/python

# First comment

print "Hello, Python!" # second comment

This produces the following result −

Hello, Python!

You can type a comment on the same line after a statement or expression −

name = "Madisetti" # This is again comment

You can comment multiple lines as follows −

# This is a comment.

# This is a comment, too.

# This is a comment, too.

# I said that already.

Following triple-quoted string is also ignored by Python interpreter and can be used as a multiline comments:

'''

This is a multiline

comment.

'''

Using Blank Lines

A line containing only whitespace, possibly with a comment, is known as a blank line and Python totally ignores it.

In an interactive interpreter session, you must enter an empty physical line to terminate a multiline statement.

Waiting for the User

The following line of the program displays the prompt, the statement saying “Press the enter key to exit”, and waits for the user to take action −

#!/usr/bin/python

raw\_input("\n\nPress the enter key to exit.")

Here, "\n\n" is used to create two new lines before displaying the actual line. Once the user presses the key, the program ends. This is a nice trick to keep a console window open until the user is done with an application.

Multiple Statements on a Single Line

The semicolon ( ; ) allows multiple statements on the single line given that neither statement starts a new code block. Here is a sample snip using the semicolon.

import sys; x = 'foo'; sys.stdout.write(x + '\n')

Multiple Statement Groups as Suites

A group of individual statements, which make a single code block are called suites in Python. Compound or complex statements, such as if, while, def, and class require a header line and a suite.

Header lines begin the statement (with the keyword) and terminate with a colon ( : ) and are followed by one or more lines which make up the suite. For example −

if expression :

suite

elif expression :

suite

else :

suite

**Command Line Arguments**

Many programs can be run to provide you with some basic information about how they should be run. Python enables you to do this with -h −

$ python -h

usage: python [option] ... [-c cmd | -m mod | file | -] [arg] ...

Options and arguments (and corresponding environment variables):

-c cmd : program passed in as string (terminates option list)

-d : debug output from parser (also PYTHONDEBUG=x)

-E : ignore environment variables (such as PYTHONPATH)

-h : print this help message and exit

You can also program your script in such a way that it should accept various options. Command Line Arguments is an advanced topic and should be studied a bit later once you have gone through rest of the Python concepts.

**Python Lists**

The list is a most versatile datatype available in Python which can be written as a list of comma-separated values (items) between square brackets. Important thing about a list is that items in a list need not be of the same type.

Creating a list is as simple as putting different comma-separated values between square brackets. For example −

list1 = ['physics', 'chemistry', 1997, 2000];

list2 = [1, 2, 3, 4, 5 ];

list3 = ["a", "b", "c", "d"]

Similar to string indices, list indices start at 0, and lists can be sliced, concatenated and so on.

A tuple is a sequence of immutable Python objects. Tuples are sequences, just like lists. The differences between tuples and lists are, the tuples cannot be changed unlike lists and tuples use parentheses, whereas lists use square brackets.

Creating a tuple is as simple as putting different comma-separated values. Optionally you can put these comma-separated values between parentheses also. For example −

tup1 = ('physics', 'chemistry', 1997, 2000);

tup2 = (1, 2, 3, 4, 5 );

tup3 = "a", "b", "c", "d";

The empty tuple is written as two parentheses containing nothing −

tup1 = ();

To write a tuple containing a single value you have to include a comma, even though there is only one value −

tup1 = (50,);

Like string indices, tuple indices start at 0, and they can be sliced, concatenated, and so on.

Accessing Values in Tuples

To access values in tuple, use the square brackets for slicing along with the index or indices to obtain value available at that index. For example −

Live Demo

#!/usr/bin/python

tup1 = ('physics', 'chemistry', 1997, 2000);

tup2 = (1, 2, 3, 4, 5, 6, 7 );

print "tup1[0]: ", tup1[0];

print "tup2[1:5]: ", tup2[1:5];

When the above code is executed, it produces the following result −

tup1[0]: physics

tup2[1:5]: [2, 3, 4, 5]

Updating Tuples

Accessing Values in Dictionary

To access dictionary elements, you can use the familiar square brackets along with the key to obtain its value. Following is a simple example −

Live Demo

#!/usr/bin/python

dict = {'Name': 'Zara', 'Age': 7, 'Class': 'First'}

print "dict['Name']: ", dict['Name']

print "dict['Age']: ", dict['Age']

When the above code is executed, it produces the following result −

dict['Name']: Zara

dict['Age']: 7

If we attempt to access a data item with a key, which is not part of the dictionary, we get an error as follows −

Live Demo

#!/usr/bin/python

dict = {'Name': 'Zara', 'Age': 7, 'Class': 'First'}

print "dict['Alice']: ", dict['Alice']

When the above code is executed, it produces the following result −

dict['Alice']:

Traceback (most recent call last):

File "test.py", line 4, in <module>

print "dict['Alice']: ", dict['Alice'];

KeyError: 'Alice'

Updating Dictionary

You can update a dictionary by adding a new entry or a key-value pair, modifying an existing entry, or deleting an existing entry as shown below in the simple example −

Live Demo

#!/usr/bin/python

dict = {'Name': 'Zara', 'Age': 7, 'Class': 'First'}

dict['Age'] = 8; # update existing entry

dict['School'] = "DPS School"; # Add new entry

print "dict['Age']: ", dict['Age']

print "dict['School']: ", dict['School']

When the above code is executed, it produces the following result −

dict['Age']: 8

dict['School']: DPS School

Delete Dictionary Elements

You can either remove individual dictionary elements or clear the entire contents of a dictionary. You can also delete entire dictionary in a single operation.

To explicitly remove an entire dictionary, just use the del statement. Following is a simple example −

Live Demo

#!/usr/bin/python

dict = {'Name': 'Zara', 'Age': 7, 'Class': 'First'}

del dict['Name']; # remove entry with key 'Name'

dict.clear(); # remove all entries in dict

del dict ; # delete entire dictionary

print "dict['Age']: ", dict['Age']

print "dict['School']: ", dict['School']

This produces the following result. Note that an exception is raised because after del dict dictionary does not exist any more −

dict['Age']:

Traceback (most recent call last):

File "test.py", line 8, in <module>

print "dict['Age']: ", dict['Age'];

TypeError: 'type' object is unsubscriptable

Note − del() method is discussed in subsequent section.

**Properties of Dictionary Keys**

Dictionary values have no restrictions. They can be any arbitrary Python object, either standard objects or user-defined objects. However, same is not true for the keys.

There are two important points to remember about dictionary keys −

(a) More than one entry per key not allowed. Which means no duplicate key is allowed. When duplicate keys encountered during assignment, the last assignment wins. For example −

Live Demo

#!/usr/bin/python

dict = {'Name': 'Zara', 'Age': 7, 'Name': 'Manni'}

print "dict['Name']: ", dict['Name']

When the above code is executed, it produces the following result −

dict['Name']: Manni

(b) Keys must be immutable. Which means you can use strings, numbers or tuples as dictionary keys but something like ['key'] is not allowed. Following is a simple example −

Live Demo

#!/usr/bin/python

dict = {['Name']: 'Zara', 'Age': 7}

print "dict['Name']: ", dict['Name']

When the above code is executed, it produces the following result −

Traceback (most recent call last):

File "test.py", line 3, in <module>

dict = {['Name']: 'Zara', 'Age': 7};

TypeError: unhashable type: 'list'

Tuples are immutable which means you cannot update or change the values of tuple elements. You are able to take portions of existing tuples to create new tuples as the following example demonstrates −

Live Demo

#!/usr/bin/python

tup1 = (12, 34.56);

tup2 = ('abc', 'xyz');

# Following action is not valid for tuples

# tup1[0] = 100;

# So let's create a new tuple as follows

tup3 = tup1 + tup2;

print tup3;

When the above code is executed, it produces the following result −

(12, 34.56, 'abc', 'xyz')

Delete Tuple Elements

Removing individual tuple elements is not possible. There is, of course, nothing wrong with putting together another tuple with the undesired elements discarded.

To explicitly remove an entire tuple, just use the del statement. For example −

Live Demo

#!/usr/bin/python

tup = ('physics', 'chemistry', 1997, 2000);

print tup;

del tup;

print "After deleting tup : ";

print tup;

This produces the following result. Note an exception raised, this is because after del tup tuple does not exist any more −

('physics', 'chemistry', 1997, 2000)

After deleting tup :

Traceback (most recent call last):

File "test.py", line 9, in <module>

print tup;

NameError: name 'tup' is not defined

**DJANGO**

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source.

Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes [reusability](https://en.wikipedia.org/wiki/Reusability)and "pluggability" of components, rapid development, and the principle of [don't repeat yourself](https://en.wikipedia.org/wiki/Don%27t_repeat_yourself). Python is used throughout, even for settings files and data models.



Django also provides an optional administrative [create, read, update and delete](https://en.wikipedia.org/wiki/Create,_read,_update_and_delete) interface that is generated dynamically through [introspection](https://en.wikipedia.org/wiki/Introspection_(computer_science)) and configured via admin models



**Create a Project**

Whether you are on Windows or Linux, just get a terminal or a cmd prompt and navigate to the place you want your project to be created, then use this code −

$ django-admin startproject myproject

This will create a "myproject" folder with the following structure −

myproject/

manage.py

myproject/

\_\_init\_\_.py

settings.py

urls.py

wsgi.py

The Project Structure

The “myproject” folder is just your project container, it actually contains two elements −

manage.py − This file is kind of your project local django-admin for interacting with your project via command line (start the development server, sync db...). To get a full list of command accessible via manage.py you can use the code −

$ python manage.py help

The “myproject” subfolder − This folder is the actual python package of your project. It contains four files −

\_\_init\_\_.py − Just for python, treat this folder as package.

settings.py − As the name indicates, your project settings.

urls.py − All links of your project and the function to call. A kind of ToC of your project.

wsgi.py − If you need to deploy your project over WSGI.

Setting Up Your Project

Your project is set up in the subfolder myproject/settings.py. Following are some important options you might need to set −

DEBUG = True

This option lets you set if your project is in debug mode or not. Debug mode lets you get more information about your project's error. Never set it to ‘True’ for a live project. However, this has to be set to ‘True’ if you want the Django light server to serve static files. Do it only in the development mode.

DATABASES = {

'default': {

'ENGINE': 'django.db.backends.sqlite3',

'NAME': 'database.sql',

'USER': '',

'PASSWORD': '',

'HOST': '',

'PORT': '',

}

}

Database is set in the ‘Database’ dictionary. The example above is for SQLite engine. As stated earlier, Django also supports −

MySQL (django.db.backends.mysql)

PostGreSQL (django.db.backends.postgresql\_psycopg2)

Oracle (django.db.backends.oracle) and NoSQL DB

MongoDB (django\_mongodb\_engine)

Before setting any new engine, make sure you have the correct db driver installed.

You can also set others options like: TIME\_ZONE, LANGUAGE\_CODE, TEMPLATE…

Now that your project is created and configured make sure it's working −

$ python manage.py runserver

You will get something like the following on running the above code −

Validating models...

0 errors found

September 03, 2015 - 11:41:50

Django version 1.6.11, using settings 'myproject.settings'

Starting development server at http://127.0.0.1:8000/

Quit the server with CONTROL-C.

A project is a sum of many applications. Every application has an objective and can be reused into another project, like the contact form on a website can be an application, and can be reused for others. See it as a module of your project.

**Create an Application**

We assume you are in your project folder. In our main “myproject” folder, the same folder then manage.py −

$ python manage.py startapp myapp

You just created myapp application and like project, Django create a “myapp” folder with the application structure −

myapp/

\_\_init\_\_.py

admin.py

models.py

tests.py

views.py

\_\_init\_\_.py − Just to make sure python handles this folder as a package.

admin.py − This file helps you make the app modifiable in the admin interface.

models.py − This is where all the application models are stored.

tests.py − This is where your unit tests are.

views.py − This is where your application views are.

Get the Project to Know About Your Application

At this stage we have our "myapp" application, now we need to register it with our Django project "myproject". To do so, update INSTALLED\_APPS tuple in the settings.py file of your project (add your app name) −

INSTALLED\_APPS = (

'django.contrib.admin',

'django.contrib.auth',

'django.contrib.contenttypes',

'django.contrib.sessions',

'django.contrib.messages',

'django.contrib.staticfiles',

'myapp',

)

Creating forms in Django, is really similar to creating a model. Here again, we just need to inherit from Django class and the class attributes will be the form fields. Let's add a forms.py file in myapp folder to contain our app forms. We will create a login form.

myapp/forms.py

#-\*- coding: utf-8 -\*-

from django import forms

class LoginForm(forms.Form):

user = forms.CharField(max\_length = 100)

password = forms.CharField(widget = forms.PasswordInput())

As seen above, the field type can take "widget" argument for html rendering; in our case, we want the password to be hidden, not displayed. Many others widget are present in Django: DateInput for dates, CheckboxInput for checkboxes, etc.

Using Form in a View

There are two kinds of HTTP requests, GET and POST. In Django, the request object passed as parameter to your view has an attribute called "method" where the type of the request is set, and all data passed via POST can be accessed via the request.POST dictionary.

Let's create a login view in our myapp/views.py −

#-\*- coding: utf-8 -\*-

from myapp.forms import LoginForm

def login(request):

username = "not logged in"

if request.method == "POST":

#Get the posted form

MyLoginForm = LoginForm(request.POST)

if MyLoginForm.is\_valid():

username = MyLoginForm.cleaned\_data['username']

else:

MyLoginForm = Loginform()

return render(request, 'loggedin.html', {"username" : username})

The view will display the result of the login form posted through the loggedin.html. To test it, we will first need the login form template. Let's call it login.html.

<html>

<body>

<form name = "form" action = "{% url "myapp.views.login" %}"

method = "POST" >{% csrf\_token %}

<div style = "max-width:470px;">

<center>

<input type = "text" style = "margin-left:20%;"

placeholder = "Identifiant" name = "username" />

</center>

</div>

<br>

<div style = "max-width:470px;">

<center>

<input type = "password" style = "margin-left:20%;"

placeholder = "password" name = "password" />

</center>

</div>

<br>

<div style = "max-width:470px;">

<center>

<button style = "border:0px; background-color:#4285F4; margin-top:8%;

height:35px; width:80%;margin-left:19%;" type = "submit"

value = "Login" >

<strong>Login</strong>

</button>

</center>

</div>

</form>

</body>

</html>

The template will display a login form and post the result to our login view above. You have probably noticed the tag in the template, which is just to prevent Cross-site Request Forgery (CSRF) attack on your site.

{% csrf\_token %}

Once we have the login template, we need the loggedin.html template that will be rendered after form treatment.

<html>

<body>

You are : <strong>{{username}}</strong>

</body>

</html>

Now, we just need our pair of URLs to get started: myapp/urls.py

from django.conf.urls import patterns, url

from django.views.generic import TemplateView

urlpatterns = patterns('myapp.views',

url(r'^connection/',TemplateView.as\_view(template\_name = 'login.html')),

url(r'^login/', 'login', name = 'login'))

When accessing "/myapp/connection", we will get the following login.html template rendered −

Setting Up Sessions

In Django, enabling session is done in your project settings.py, by adding some lines to the MIDDLEWARE\_CLASSES and the INSTALLED\_APPS options. This should be done while creating the project, but it's always good to know, so MIDDLEWARE\_CLASSES should have −

'django.contrib.sessions.middleware.SessionMiddleware'

And INSTALLED\_APPS should have −

'django.contrib.sessions'

By default, Django saves session information in database (django\_session table or collection), but you can configure the engine to store information using other ways like: in file or in cache.

When session is enabled, every request (first argument of any view in Django) has a session (dict) attribute.

Let's create a simple sample to see how to create and save sessions. We have built a simple login system before (see Django form processing chapter and Django Cookies Handling chapter). Let us save the username in a cookie so, if not signed out, when accessing our login page you won’t see the login form. Basically, let's make our login system we used in Django Cookies handling more secure, by saving cookies server side.

For this, first lets change our login view to save our username cookie server side −

def login(request):

username = 'not logged in'

if request.method == 'POST':

MyLoginForm = LoginForm(request.POST)

if MyLoginForm.is\_valid():

username = MyLoginForm.cleaned\_data['username']

request.session['username'] = username

else:

MyLoginForm = LoginForm()

return render(request, 'loggedin.html', {"username" : username}

Then let us create formView view for the login form, where we won’t display the form if cookie is set −

def formView(request):

if request.session.has\_key('username'):

username = request.session['username']

return render(request, 'loggedin.html', {"username" : username})

else:

return render(request, 'login.html', {})

Now let us change the url.py file to change the url so it pairs with our new view −

from django.conf.urls import patterns, url

from django.views.generic import TemplateView

urlpatterns = patterns('myapp.views',

url(r'^connection/','formView', name = 'loginform'),

url(r'^login/', 'login', name = 'login'))

When accessing /myapp/connection, you will get to see the following page

**CONCLUSION**

The quality of fundus images is crucial for ensuring the diagnostic reliability of the ophthalmologist or automated medical system. To enhance the DR grading performance on the low-quality fundus images, we propose an end-to-end quality assessment guided collaborative learning framework that (1) improves the disease grading performance given a large number of low-quality images, (2) achieves fundus image quality enhancement, and (3) trains an image quality assessment model. The experimental results demonstrate that our method significantly improves the latest results of DR grading on benchmark fundus datasets, and the low-quality fundus images also gain remarkable enhancement

**REFERENCE**

[1] K. Ogurtsova et al., “IDF diabetes atlas: Global estimates of undiagnosed diabetes in adults for 2021,” Diabetes Res. Clin. Pract., vol. 183, 2022, Art. no. 109118.

[2] V. Mayya, S. Kamath, and U. Kulkarni, “Automated microaneurysms detection for early diagnosis of diabetic retinopathy: A comprehensive review,” Comput. Methods Prog. Biomed. Update, vol. 1, 2021, Art. no. 100013.

[3] Y. Zhou, B. Wang, L. Huang, S. Cui, and L. Shao, “A benchmark for studying diabetic retinopathy: Segmentation, grading, and transferability,” IEEE Trans. Med. Imag., vol. 40, no. 3, pp. 818–828, Mar. 2021.

[4] T. Köhler, A. Budai, M. F. Kraus, J. Odstrˇcilik, G. Michelson, and J. Hornegger, “Automatic no-reference quality assessment for retinal fundus images using vessel segmentation,” inProc. 26th IEEE Int. Symp. Comput.- Based Med. Syst., 2013, pp. 95–100.

[5] S. Wang, K. Jin, H. Lu, C. Cheng, J. Ye, and D. Qian, “Human visual system-based fundus image quality assessment of portable fundus camera photographs,” IEEE Trans. Med. Imag., val. 35, no. 4, pp. 1046–1055, Apr. 2016.

[6] F. Yu, J. Sun, A. Li, J. Cheng, C. Wan, and J. Liu, “Image quality classification for DR screening using deep learning,” in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2017, pp. 664–667.

[7] G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. T. Salles, “Retinal image quality assessment using deep learning,” Comput. Biol. Med., vol. 103, pp. 64–70, 2018.

[8] C. Lei and Q. Chen, “Robust reflection removal with reflection-free flashonly cues,” inProc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 14811–14820.

[9] Y. Ye, Y. Chang, H. Zhou, and L. Yan, “Closing the loop: Joint rain generation and removal via disentangled image translation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 2053–2062.

[10] S. J. Cho, S. W. Ji, J. P. Hong, S. W. Jung, and S. J. Ko, “Rethinking coarse-to-fine approach in single image deblurring,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 4641–4650.

[11] E. Decencière et al., “Feedback on a publicly distributed image database: The messidor database,” Image Anal. Stereol., vol. 33, no. 3, pp. 231–234, 2014.

[12] S. Liu, Y. Liang, and A. Gitter, “Loss-balanced task weighting to reduce negative transfer in multi-task learning,” in Proc. AAAI Conf. Artif. Intell., 2019, pp. 9977–9978.

[13] Y. Ganin and L. Victor, “Unsupervised domain adaptation by backpropagation,” in Proc. Int. Conf. Mach. Learn., 2015, pp. 1180–1189.

[14] H. Fu et al., “Evaluation of retinal image quality assessment networks in different color-spaces,” in Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv., Springer, 2019, pp. 48–56.

[15] “Kaggle diabetic retinopathy detection competition.” [Online]. Available: <https://www.kaggle.com/c/diabetic-retinopathy-detection>

[16] J. Staal, M. D. Abramoffff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” IEEE Trans. Med. Imag., vol. 23, no. 4, pp. 501–509, Apr. 2004.

[17] Z. Gu et al., “CE-Net: Context encoder network for 2D medical image segmentation,” IEEE Trans. Med. Imag., vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[19] Kaggle diabetic retinopathy detection competition. [Online]. Available: <https://www.kaggle.com/c/diabetic-retinopathy-detection/leaderboard>

[20] X. Li, X. Hu, L. Yu, L. Zhu, C. W. Fu, and P. A. Heng, “CANet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading,” IEEE Trans. Med. Imag., vol. 39, no. 5, pp. 1483–1493, May 2020.