COMP9444: Neural Networks and Deep Learning
Project Description List

Content

Project ID: 001

Project Title: Insect pest specifies identification.

Area of Research: Computer Vision

Problem Statement: Insect pest classification plays a crucial role in various domains, including agriculture, pest control, and ecological research. Rapid and accurate identification of insect pests is essential for effective pest management strategies, early detection of invasive species, and preservation of crop yield and quality. However, manual classification of insects based on visual inspection can be time-consuming, error-prone, and challenging, particularly when dealing with conditions in the wild. The goal of this project is to correctly identify the species of insects in an automated manner using advanced artificial intelligence algorithms which has high accuracy, robust to varying environmental conditions, appearance, and deploying these algorithms for real-time monitoring.

Dataset:

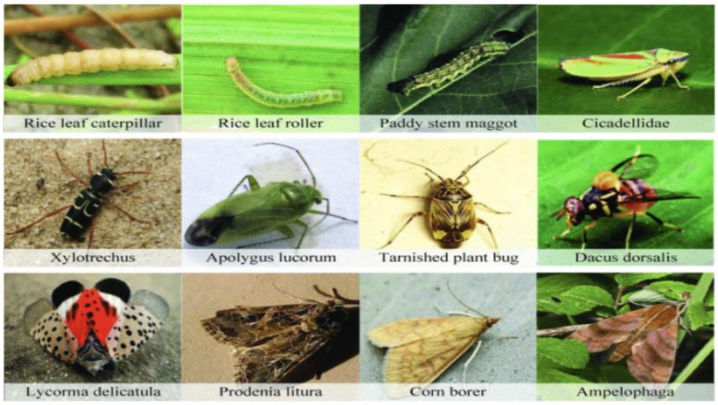


Figure 1. Example images of the IP102 dataset. Each image belongs to a different species of insect pests.

The IP102 dataset [1] is a benchmark dataset for insect pests. The details about the dataset can be found in [1] and be downloaded from the URL given below:

Dataset URL: https://github.com/xpwu95/IP102/tree/master

Task: To develop an automatic insect recognition system using neural networks and deep learning which provides high accuracy, robust to varying appearance and similarity between different insect species, and faster so that it can be deployed in the real world.

Relevant Papers

[1]. X. Wu, C. Zhan, Y. -K. Lai, M. -M. Cheng and J. Yang, "IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 8779-8788, doi: 10.1109/CVPR.2019.00899.
https://ieeexplore.ieee.org/document/8954351
https://openaccess.thecvf.com/content_CVPR_2019/papers/Wu_IP102_A_Large-Scale_Benchmark_Dataset_for_Insect_Pest_Recognition_CVPR_2019_paper.pdf

[2]. A. Setiawan, N Yudistira, and R.C. Wihandika, “Large scale pest classification using efficient Convolutional Neural Network with augmentation and regularizers”, Computers and Electronics in Agriculture, Vol. 200, Sept 2022.
https://www.sciencedirect.com/science/article/pii/S0168169922005191

[3]. W. Linfeng, L. Yong, L. Jiayao, W. Yunsheng, and X. Shipu, “Based on the multi-scale information sharing network of fine-grained attention for agricultural pest detection”, PLOS ONE 18(10):e0286732.
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0286732

[4]. An J, Du Y, Hong P, Zhang L, Weng X., “Insect recognition based on complementary features from multiple views”, Scientific Reports. 2023 Feb;13(1):2966. DOI: 10.1038/s41598-023-29600-1.
https://europepmc.org/article/pmc/pmc9940688

[5]. S. Kar, J. Nagasubramanian, D. Elango, M. E. Carroll, C. A. Abel, A. Nair, D.S. Mueller, M. E. O’Neal, A. K. Singh, S. Sarkar, B. Ganapathysubramanian, A. Singh, “Self-supervised learning improves classification of agriculturally important insect pests in plants”, The Plant Phenome Journal, 6, e20079.
https://acsess.onlinelibrary.wiley.com/doi/full/10.1002/ppj2.20079

Project ID: 002

Project Title: Skin Lesion Classification Using Deep Learning

Area of Research: Computer Vision (CV)

Problem Statement:

The project aims to develop a deep learning model for accurately classifying skin lesions into distinct categories, including melanoma, nevi, and benign lesions. Skin lesion classification is a critical task in dermatology, and automating this process using deep learning techniques can aid in early detection and improved patient outcomes. Unlike standard datasets such as MNIST or CIFAR-10, which are limited in diversity, skin lesion classification presents challenges due to the variability and complexity of skin conditions.

Dataset:

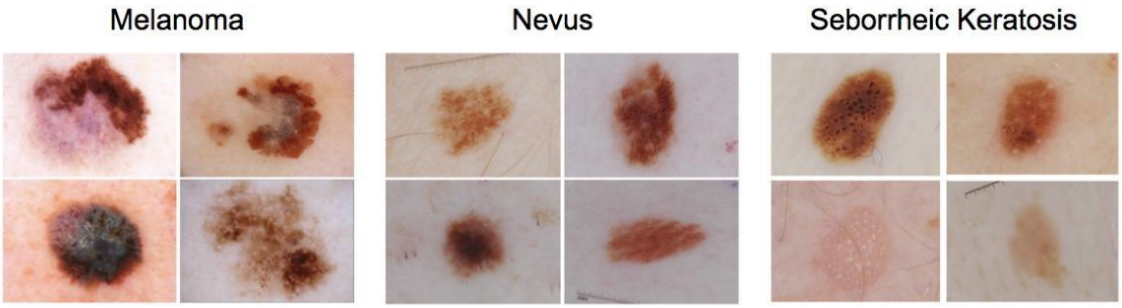


Figure 1: example images of ISIC dataset.

The project utilizes the International Skin Imaging Collaboration (ISIC) dataset, which contains a diverse collection of dermatoscopic images encompassing various skin lesion types. The dataset is annotated with ground truth labels for lesion categories, providing a valuable resource for supervised learning tasks in skin lesion classification.

Dataset URL: https://challenge.isic-archive.com/data/

Task:

The task involves training a deep neural network to accurately classify skin lesions into predefined categories, including melanoma (malignant), nevi, and benign lesions. Students will explore architectures such as ResNet, VGG, or custom networks suitable for handling complex image features specific to dermatoscopic images. Data augmentation techniques such as rotation, flipping, scaling, and color jittering will be employed to improve model generalization and robustness.

Relevant Papers:

1. Kassem, Mohamed A., et al. "Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review." Diagnostics 11.8 (2021): 1390.

2. Lopez, Adria Romero, et al. "Skin lesion classification from dermoscopic images using deep learning techniques." 2017 13th IASTED international conference on biomedical engineering (BioMed). IEEE, 2017.

3. Benyahia, Samia, Boudjelal Meftah, and Olivier Lézoray. "Multi-features extraction based on deep learning for skin lesion classification." Tissue and Cell 74 (2022): 101701.

Project ID: 003

Project Title: NLP Classification of FOMC Texts for Market Analysis

Area of Research: Natural Language Processing (NLP)

Problem Statement:
In modern financial markets, the Federal Reserve's (Fed) monetary policy announcements, especially the Federal Open Market Committee (FOMC) statements, significantly influence market trends. Consequently, there is an urgent need for a model capable of extracting policy stances from FOMC texts to better understand and predict their impact on financial markets. This project aims to leverage advanced artificial intelligence techniques to classify sentences in FOMC texts as hawkish, dovish, or neutral. This provides a new tool for financial market analysis, offering market participants a more accurate means of assessing policy stances and aiding in the better understanding and prediction of the FOMC statements' effects on financial markets.

Dataset:

	A	B	C	D
1	index	sentence	year	label
2	487	Instead,	1998	2
3	422	Inflatior	2004	0
4	150	It isn’ t	2021	2
5	110	At some p	2022	1
6	548	Moreover,	2002	0
7	122	Based on	2007	0
8	282	We have l	2020	2

Figure 1. This figure is a partial display of the FOMC text dataset, which spans from 1996 to 2022. The dataset includes text records from FOMC speeches, meeting minutes, and press conferences.

Regarding the meeting minutes, there are 214 documents containing 20,618 target sentences. The press conferences comprise a total of 63 documents with 5,086 target sentences. For speeches, there are 1,026 documents (filtered down to 201), with 12,465 target sentences. The specific links are as follows:

Dataset URL: https://github.com/gtfintechlab/fomc-hawkish-dovish/tree/main/training_data/test-and-training

Dataset paper: https://arxiv.org/pdf/2305.07972

Task: This project proposes the task of classifying Federal Open Market Committee (FOMC) texts. By implementing a "hawkish-dovish" classification task, it allows for a more accurate extraction of monetary policy stances from FOMC texts. This detailed classification aids in understanding and predicting the specific impacts of FOMC statements on financial markets, thereby providing market participants and policymakers with more accurate and valuable analytical tools.

Relevant Papers

1. Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10 - Ks[J]. The Journal of finance, 2011, 66(1): 35-65. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2010.01625.x

2. Hansen S, McMahon M. Shocking language: Understanding the macroeconomic effects of central bank communication[J]. Journal of International Economics, 2016, 99: S114-S133. https://www.sciencedirect.com/science/article/pii/S0022199615001828

3. Araci D. Finbert: Financial sentiment analysis with pre-trained language models[J]. arXiv preprint arXiv:1908.10063, 2019. https://arxiv.org/pdf/1908.10063

4. Mathur P, Neerkaje A, Chhibber M, et al. Monopoly: Financial prediction from monetary policy conference videos using multimodal cues[C]//Proceedings of the 30th ACM International Conference on Multimedia. 2022: 2276-2285. https://dl.acm.org/doi/abs/10.1145/3503161.3548380

5. Shah A, Paturi S, Chava S. Trillion dollar words: A new financial dataset, task & market analysis[J]. arXiv preprint arXiv:2305.07972, 2023. https://arxiv.org/pdf/2305.07972

Project ID: 004

Project Title: Sentiment Analysis from Video, Audio, and Text Data

Area of Research: Multimodal AI

Problem Statement: Sentiment analysis is a critical task in natural language processing and machine learning, aimed at determining the sentiment expressed in each piece of text. Traditional approaches to sentiment analysis predominantly focus on text data alone. However, human communication is inherently multimodal, encompassing not just words, but also tone of voice, facial expressions, and body language. These additional modalities can provide valuable context and improve the accuracy and depth of sentiment analysis.

The CMU-MOSEI dataset offers an opportunity for students to explore sentiment analysis from a multimodal perspective by providing synchronized video, audio, and text data. Each sample in the dataset is a video segment where an individual expresses their opinion on a topic. These video segments are accompanied by audio tracks and text transcripts, with each modality annotated for sentiment intensity.

More details of the dataset are provided in Dataset URL.

Dataset:

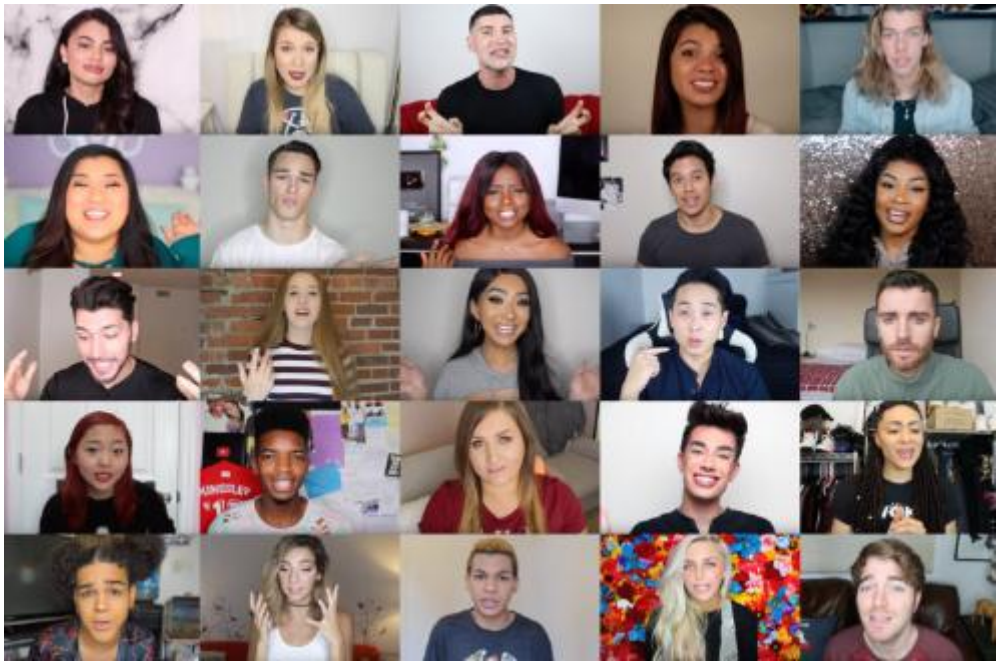


Figure 1. Screenshot examples of video segments from various Youtubers in CMU-MOSEI dataset.

Dataset URL: https://github.com/CMU-MultiComp-Lab/CMU-MultimodalSDK

Task: To develop a machine learning model capable of analyzing sentiment from multimodal data, including video, audio, and text. This project aims to combine these different modalities to achieve a more accurate sentiment classification.

Relevant Papers

[1]. Zadeh, A., Liang, P.P., Mazumder, N., Poria, S., Cambria, E. and Morency, L.P., 2018, April. Memory fusion network for multi-view sequential learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No.1). https://ojs.aaai.org/index.php/AAAI/article/view/12021

[2]. Nojavanasghari, B., Baltrušaitis, T., Hughes, C.E. and Morency, L.P., 2016, October. Emoreact: a multimodal approach and dataset for recognizing emotional responses in children. In Proceedings of the 18th acm international conference on multimodal interaction (pp. 137-144). https://dl.acm.org/doi/abs/10.1145/2993148.2993168

Project ID: 005

Project Title: Intelligent Food Image Recognition

Area of Research: Computer Vision

Problem Statement:
People are becoming increasingly aware of the importance of diet control, recognizing its role in analyzing and maintaining long-term health. Diet directly impacts daily physical activity and plays a crucial role in disease prevention and recovery. Therefore, tracking and monitoring daily food intake is highly valuable. Given the challenges of food image recognition and analysis, this project employs advanced deep learning methods to achieve high-precision segmentation of food images. This allows users to understand their eating habits and set and achieve health goals.

Dataset:

Figure 1. Example image showcases the UECFoodPix dataset, with each image displaying various foods along with their corresponding annotations.

UECFoodPix [1] is a food image dataset that includes segmentation masks. The detailed link and paper are provided below:

Dataset URL: https://mm.cs.uec.ac.jp/uecfoodpix/

Dataset paper: https://link.springer.com/chapter/10.1007/978-3-030-68821-9_51

Task: Develop an automated food image recognition system using neural networks and deep learning. The system should maintain high-precision segmentation. It aims to help users understand their eating habits and achieve health goals. The system's efficiency and accuracy make it suitable for real-world deployment and application.

Relevant Papers

[1]. A. Myers, et al., "Im2Calories: Towards an Automated Mobile Vision Food Diary," in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015 pp. 1233-1241. doi: 10.1109/ICCV.2015.146
https://www.computer.org/csdl/proceedings-article/iccv/2015/8391b233/12OmNvSbBJ8

[2]. C. N. C. Freitas, F. R. Cordeiro and V. Macario, "MyFood: A Food Segmentation and Classification System to Aid Nutritional Monitoring," 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Porto de Galinhas, Brazil, 2020, pp. 234-239, doi: 10.1109/SIBGRAPI51738.2020.00039.
https://ieeexplore.ieee.org/abstract/document/9265999

[3]. Dai, Y.; Park, S.; Lee, K. Utilizing Mask R-CNN for Solid-Volume Food Instance Segmentation and Calorie Estimation. Appl. Sci. 2022, 12, 10938. https://doi.org/10.3390/app122110938
https://www.mdpi.com/2076-3417/12/21/10938

[4]. Kong X Y, Sun X H, Wang Y Z, et al. Food Calorie Estimation System Based on Semantic Segmentation Network[J]. Sensors & Materials, 2023, 35.
https://sensors.myu-group.co.jp/sm_pdf/SM3304.pdf

[5]. Dong X, Li H, Wang X, et al. CANet: cross attention network for food image segmentation[J]. Multimedia Tools and Applications, 2023: 1-20.
https://link.springer.com/article/10.1007/s11042-023-17916-z

Project ID: 006

Project Title: Post-Earthquake Building Damage Detection

Area of Research: Computer Vision

Problem Statement: The impact of earthquakes is extensive, and these sudden events can simultaneously cause multiple structural defects in buildings. Traditional manual methods struggle to effectively track and identify these issues, especially in large-scale disasters. Quickly and accurately identifying post-earthquake building damage is crucial for developing rescue strategies, ensuring structural safety, and assessing the extent of the damage. This project aims to leverage advanced artificial intelligence algorithms to accurately detect and classify earthquake-induced building defects. The algorithms must demonstrate high accuracy and adaptability to various environmental conditions and appearances, ultimately providing reliable technical support for post-disaster recovery.

Dataset:

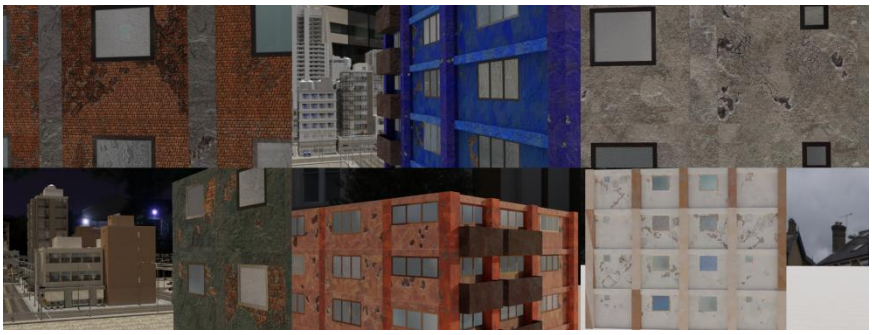


Figure 1. Sample images from the QuakeCity dataset, each with six annotations: damage masks (cracks, spalling, exposed rebar), component, component damage status, and depth map.

The QuakeCity dataset [1], generated using physics-based graphical models (PBGM), was introduced by Hoskere et al. in 2021. The detailed information can be found at the following link:

Dataset URL: https://sail.cive.uh.edu/quakecity/

Task: Develop an automated earthquake damage detection system for buildings using deep learning and neural network technology. The system should maintains high accuracy and speed in capturing building defects under various environmental conditions and appearance changes, providing scientific and effective support for post-disaster recovery.

Relevant Papers

[1]. V. Hoskere, “Developing autonomy in structural inspections through computer vision and graphics,” University of Illinois at Urbana-Champaign, 2021.

https://www.ideals.illinois.edu/items/117278

[2]. M. Meng, K. Zhu, K. Chen, and H. Qu, “A modified fully convolutional network for crack damage identification compared with conventional methods,” Modelling and Simulation in Engineering, vol. 2021, pp. 1 – 14, 11 2021 https://www.hindawi.com/journals/mse/2021/5298882/

[3]. V. Hoskere, Y. Narazaki, B. F. Spencer Jr., Physics-Based Graphics Models in 3D Synthetic Environments as Autonomous Vision-Based Inspection Testbeds. Sensors 2022, 22, 532.

https://doi.org/10.3390/s22020532

[4]. M. Żarski, B. Wójcik, J. A. Miszczak, B. Blachowski and M. Ostrowski, "Computer Vision Based Inspection on Post-Earthquake With UAV Synthetic Dataset," in IEEE Access, vol. 10, pp. 108134-108144, 2022, doi: 10.1109/ACCESS.2022.3212918.’

https://ieeexplore.ieee.org/abstract/document/9913977

Project ID: 007

Project Title: Abstractive Document Summarization.

Area of Research: Natural Language Processing.

Problem Statement: Automatic text summarization involves a process of shortening a source text efficiently while keeping the main idea intact, which aids in reducing the amount of time required to process information, helps with faster search for information, and makes learning one topic easier. In general, automatic text summarization can be conceptualized as having three approaches: extractive, abstractive, and hybrid approach. This project will mainly focus on the abstractive approach, that comprehends a source document and writes a summary output based on the salient concepts of the source document, imitating a human.

Dataset:

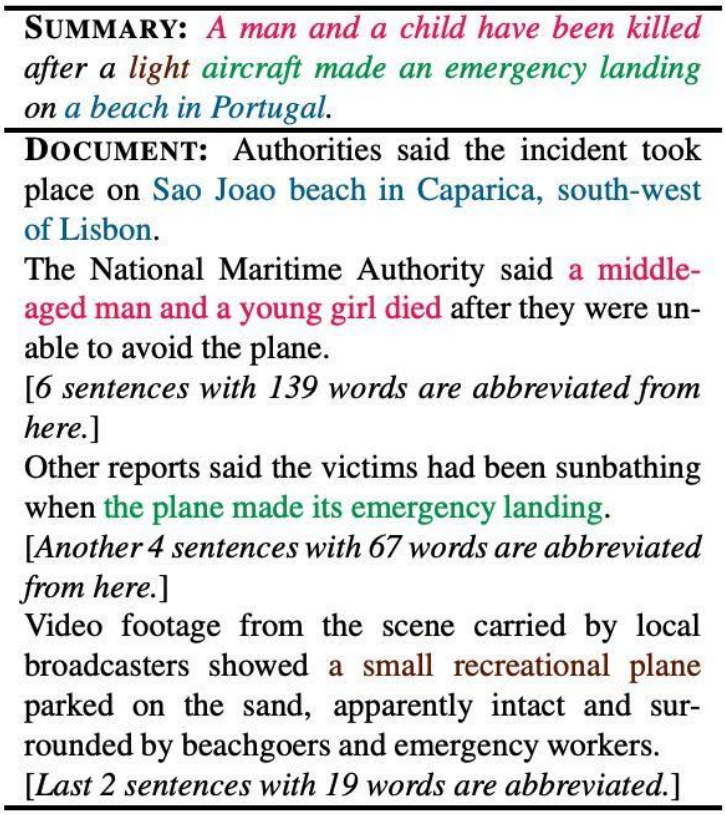


Figure 1. An abridged example from extreme summarisation dataset showing the document and its one line summary. Document content present in the summary is color-coded.

XSum[1] dataset consists of BBC articles and accompanying single sentence summaries.

Dataset URL: https://huggingface.co/datasets/EdinburghNLP/xsum

Dataset paper: https://aclanthology.org/D18-1206

Task: The goal of this project is to perform abstractive document summarisation using advanced artificial intelligence algorithms that yield increased performance on relevant metrics while making sure to generate accurate and relevant summaries.

Relevant Papers

[1]. Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics. https://aclanthology.org/D18-1206

[2]. Huan Yee Koh, Jiaxin Ju, Ming Liu, and Shirui Pan. 2022. An Empirical Survey on Long Document Summarization: Datasets, Models, and Metrics. ACM Comput. Surv. 55, 8, Article 154 (August 2023), 35 pages. https://doi.org/10.48550/arXiv.2207.00939

Project ID: 008

Project Title: Custom Chatbots with LLMs

Area of Research: LLMs, Langchain, RAG, VectorDB

Problem Statement: Large Language Models (LLMs) have revolutionized the field of natural language processing by enabling machines to understand and generate human-like text with remarkable accuracy. Despite these advancements, developing a customized chatbot that can effectively leverage LLMs for specific domains remains challenging. LLMs need to be fine-tuned to understand domain-specific terminology, context, and nuances to provide accurate and relevant responses. Additionally, integrating LLMs with vector databases can enhance the retrieval and contextual relevance of information. The goal of this project is to leverage the power of LangChain VectorDB, LLM to create a customizable chatbot that excels in understanding and responding to specialized domain queries.

Dataset:

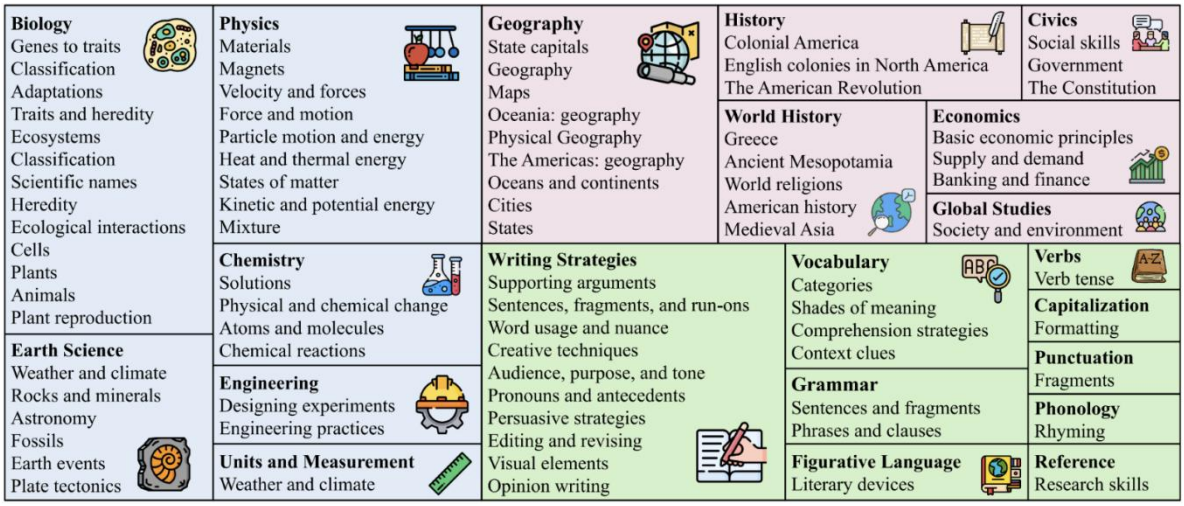


Figure 1. ScienceQA (Science Question Answering). For more details, you can explore the datatset and check the visualizations here: https://scienceqa.github.io/explore.html

Science Question Answering (ScienceQA) is a new benchmark collected from elementary and high school science curricula, containing 21,208 multimodal multiple-choice science questions. It consists of a diverse set of science topics and annotations of their answers with corresponding lectures and explanations. It further designs language models to learn to generate lectures and explanations as the chain of thought (CoT) to mimic the multi-hop reasoning process when answering ScienceQA questions.

LLM Base-model: GPT-4
Dataset URL: https://scienceqa.github.io/#dataset

Dataset paper: https://arxiv.org/pdf/2209.09513

Task: To develop a sophisticated chatbot using LangChain VectorDB LLM that can accurately understand and respond to a wide range of domain-specific user queries, integrate seamlessly with vector databases to improve the relevancy and accuracy of responses. The custom chatbot shall be able to handle various types of questions, including open-ended questions, multiple-choice questions, and context-based queries.

Relevant Papers

[1]. Lu, P., Mishra, S., Xia, T., Qiu, L., Chang, K.W., Zhu, S.C., Tafjord, O., Clark, P. and Kalyan, A., 2022. Learn to explain: Multimodal reasoning via thought chains for science question answering. *Advances in Neural Information Processing Systems*, *35*, pp.2507-2521.
https://arxiv.org/pdf/2209.09513
[2] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, *33*, pp.9459-9474.
https://arxiv.org/pdf/2005.11401
[3] Langchain github: https://python.langchain.com/v0.2/docs/introduction/
https://github.com/langchain-ai/langchain

Project ID: 009

Project Title: Maize Nitrogen Deficiency Detection and Classification

Area of Research: Computer Vision

Problem Statement:
Nitrogen fertilization is one of the most expensive inputs in maize production. Additionally, nitrate and nitrite leakage is known to cause various medical conditions and damage natural ecosystems. Therefore, this project focuses on extracting agronomic indices of maize under different nitrogen fertilization scenarios using machine learning techniques on RGB images. The tasks of detection and classification are crucial in this process, as accurate detection and classification of nitrogen deficiency can help optimize fertilization strategies. The goal of this effort is to optimize nitrogen usage on farmers' fields, thereby reducing its environmental footprint.

Dataset:

Figure 1. The dataset shown in Figure 1 is titled "Nitrogen Deficiency in Maize." Each image in this dataset displays plants under three different nitrogen application levels: N0, N75, and NFull.

This dataset is intended for classification tasks and has a size of 489MB. The detailed link is provided below:

Dataset URL: https://data.mendeley.com/datasets/g7xnn2bm4g/1

Task: Develop a system to detect and classify nitrogen deficiency in maize using Artificial intelligence on RGB images. The system will leverage advanced algorithms to accurately extract agronomic indices, aiding in the optimization of nitrogen fertilization strategies. This approach aims to enhance precision in applying nitrogen, reducing both costs and environmental impact, making it highly applicable for real-world agricultural deployment.

Relevant Papers

[1]. A.D.J. van Dijk, G. Kootstra, W. Kruijer, D. de Ridder, Machine learning in plant science and plant breeding, IScience 24 (2021) 101890, doi:10.1016/j.isci.2020.101890.
https://www.cell.com/iscience/pdf/S2589-0042(20)31087-7.pdf

[2]. H. Tong, Z. Nikoloski, Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data, J. Plant Physiol. 257 (2021) 153354, doi:10.1016/j.jplph.2020.153354
https://www.sciencedirect.com/science/article/pii/S0176161720302443

[3]. J.S. Spišic, ´ D. Šimic, ´ J. Balen, A.J. Jambrovic, ´ V. Galic, ´ Machine Learning in the Analysis of Multispectral Reads in Maize Canopies Responding to Increased Temperatures and Water Deficit, 2022, p. 2596, doi:10.3390/rs14112596.
https://www.mdpi.com/2072-4292/14/11/2596

[4]. Salaić M, Novoselnik F, Žarko I P, et al. Nitrogen deficiency in maize: Annotated image classification dataset[J]. Data in brief, 2023, 50: 109625.
https://www-sciencedirect-com.wwwproxy1.library.unsw.edu.au/science/article/pii/S2352340923007102?via%3Dihub

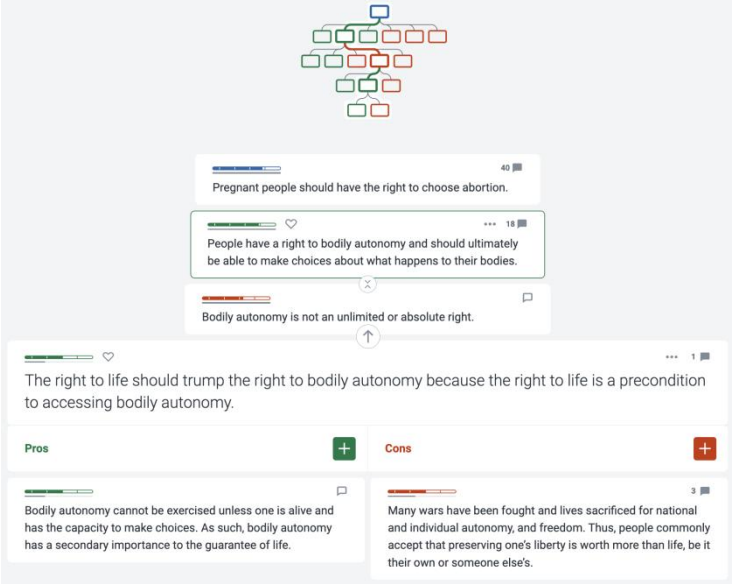
Project ID: 010

Project Title: Argument Move Classification

Area of Research: Natural Language Processing (NLP).

Problem Statement: Argumentative Dialogues are conversations involving some form of opinionated content (e.g. politics, social issues etc), in which the users produce arguments to convince each other of their point of view. A crucial component of understanding these conversations is being able to correctly understand which Move a user is performing (e.g. whether they are attacking or supporting another point that has been made in the discussion). The goal of this project is to build a neural architecture that is able to correctly classify the Argumentative Move of a user given the history of the dialogue so far.

Dataset:



Kialo (https://www.kialo.com) is an online platform for debates that provides Argumentative Move annotation for each contribution by the users. A dataset containing debates from the website can be downloaded from the URL below

Dataset URL: https://netsys.surrey.ac.uk/datasets/graphnli/orig_data/

Task: Classify posts from an online forum debate based on whether they are ***Attacking*** or ***Supporting*** the previous post in the debate

Relevant Papers

[1]. Agarwal, Vibhor, et al. "GraphNLI: A graph-based natural language inference model for polarity prediction in online debates." Proceedings of the ACM Web Conference 2022. 2022. https://dl.acm.org/doi/abs/10.1145/3485447.3512144?casa_token=R5vpZ5Z8k6kAAAAA:Zq8sIqvwjbDdgPIxbAPVCMH1KHFuNI2u0rWYIXm58ImfUs3gnoV73R6x3UxjePUaf4yLeuS7S2n5

[2]. Jo, Yohan, et al. "Classifying argumentative relations using logical mechanisms and argumentation schemes." Transactions of the Association for Computational Linguistics 9 (2021): 721-739. https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00394/106790/Classifying-Argumentative-Relations-Using-Logical

Project ID: 011

Project Title: Hepatic Vessels Segmentation Using CT Images

Area of Research: Computer Vision (CV)

Problem Statement: The project aims to develop a deep learning model for segmenting hepatic vessels from CT (Computed Tomography) images. Accurate segmentation of liver vessels is crucial for surgical planning, disease diagnosis, and medical research. This project addresses the challenge of automating hepatic vessel segmentation to improve clinical workflows and patient outcomes.

Dataset: Decathlon challenge dataset

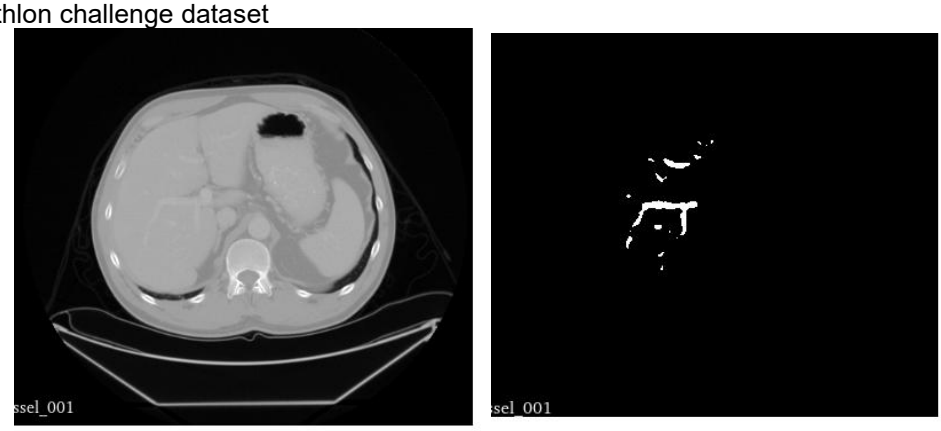


Figure 1:sample CT image and Hepatic vessels ground truth.

Dataset URL: https://drive.google.com/drive/folders/1HqEgzS8BV2c7xYNrZdEAnrHk7osJJ--2

Dataset Paper:
Antonelli, Michela, et al. "The medical segmentation decathlon." Nature communications 13.1 (2022): 4128.
Task: The task involves semantic segmentation to delineate hepatic vessels from CT scans. Students will explore deep learning techniques such as convolutional neural networks (CNNs), U-Net architectures, and data augmentation methods for accurate vessel segmentation.

Relevant Papers:
1. Alirr, Omar Ibrahim, and Ashrani Aizzuddin Abd Rahni. "Hepatic vessels segmentation using deep learning and preprocessing enhancement." Journal of applied clinical medical physics 24.5 (2023): e13966.

2. Lebre, Marie-Ange, et al. "Automatic segmentation methods for liver and hepatic vessels from CT and MRI volumes, applied to the Couinaud scheme." Computers in biology and medicine 110 (2019): 42-51.

3. Kitrungrotsakul, Titinunt, et al. "VesselNet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation." Computerized Medical Imaging and Graphics 75 (2019): 74-83.

Project ID: 012

Project Title: Autonomous Drone Navigation using Deep Reinforcement Learning.

Area of Research: Reinforcement Learning and Deep Learning.

Problem Statement: Autonomous drone navigation is a complex task that involves planning and decision-making in dynamic and uncertain environments to successfully navigate a drone. Traditional methods often struggle with adapting to changes in the environment, such as obstacles. This project aims to develop a deep reinforcement learning-based system for autonomous drone navigation that combines reinforcement learning and deep learning approaches. The system will enable a drone to navigate through various environments, avoid obstacles, and reach a designated target efficiently. The primary objective is to train a drone to navigate using visual inputs, making real-time decisions based on its surroundings. For this project, we recommend using the AirSim simulation platform provided by Microsoft research team, which provides a realistic and configurable environment for training and testing autonomous drones.

Dataset:

Figure 1. An example environment and scenario with a drone within the AirSim simulator. For watching the whole video, check this link: https://www.youtube.com/watch?v=-WfTr1-OBGQ

Dataset URL: https://github.com/microsoft/AirSim

Dataset paper: Shah, Shital, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. "AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles." Field and Service Robotics. Springer, Cham, 2018. 621-635.

Task: Firstly, you should be able to set up AirSim simulator and configure various environments for training the drone. This includes creating diverse scenarios with different obstacles, visual inputs from the drone's camera, position, and velocity. Then, you need to develop a reinforcement learning algorithm to train the drone agent. The algorithm should enable the drone to learn optimal navigation policies from its interactions with the environment. The trained agent in various test scenarios must navigate efficiency, avoid obstacle, and have an acceptable task completion rate.

Relevant Papers:

[1]. Shah, Shital, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. "AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles." Field and Service Robotics. Springer, Cham, 2018. 621-635.

[2]. Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529-533.

[3]. Schulman, John, et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).

Project ID: 013

Project Title: Mosquitos on Human Skin Recognition

Area of Research: Computer Vision

Problem Statement: Mosquitoes pose substantial threat to public health resulting in million number of deaths worldwide every year. They act as the vectors responsible for diseases such as Dengue, Yellow fever, Chikungunya, Zika etc. The harmful mosquito species are contained in the genera Aedes, Anopheles and Culex. Since the differences between species are trivial for human vision, the current prevailing approach for identifying mosquito's species is based on recording the frequency of wing flapping, image classification of mosquitos is still a developing problem that requiring contributions. The goal of this project is to offer a chance for students to solve an image classification problem that outside of toy datasets such as MNIST and CIFAR10, using algorithms learned in lectures, at the same time leading them to a practical problem that strong connected to making contributions for human health.
Details of the dataset are provided in Dataset URL.

Dataset:

Figure 1. Examples of mosquitos in six different classes. The species number are three, and there are two conditions for each instance (landed or smashed).

Dataset URL: https://data.mendeley.com/datasets/zw4p9kj6nt/2

Task: Develop a mosquito classification model using neural networks and deep learning algorithms that obtain a performance on the test set with accuracy as higher as you can.

Relevant Papers

[1]. Ong, S.Q. and Ahmad, H., 2022. An annotated image dataset for training mosquito species recognition system on human skin. Scientific Data, 9(1), p.413.
https://www.nature.com/articles/s41597-022-01541-w

[2]. Park, J., Kim, D.I., Choi, B., Kang, W. and Kwon, H.W., 2020. Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. Sci Rep 10: 1012.
https://www.nature.com/articles/s41598-020-57875-1

[3]. Adhane, G., Dehshibi, M.M. and Masip, D., 2021. A deep convolutional neural network for classification of aedes albopictus mosquitoes. IEEE Access, 9, pp.72681-72690.
https://ieeexplore.ieee.org/abstract/document/9429188

Project ID: 014

Project Title: Automated Yoga Pose Classification

Area of Research: Computer Vision

Problem Statement: Yoga pose classification plays a crucial role in various domains, including fitness, physical therapy, and sports training. Rapid and accurate identification of yoga poses is essential for effective practice monitoring, correction of postures, and personalized training programs. However, manual classification of yoga poses based on visual inspection can be time-consuming, error-prone, and challenging, particularly when dealing with a wide variety of poses and subtle differences between them. The goal of this project is to correctly identify different yoga poses in an automated manner using neutral network, achieving high accuracy, robustness to varying appearances and angles, and real-time deployment capabilities.

Dataset:

Figure 1. Example images of Yoga-82 dataset, each line represents a type of pose of yoga.

Yoga-82 dataset [1] is a benchmark dataset for yoga pose detection. The details about the dataset can be found in [1] and be downloaded from the URL given below:
Dataset URL: https://neurohive.io/en/news/yoga-82-new-dataset-with-complex-yoga-poses/

Task: To develop an automatic yoga pose recognition system using neural networks and deep learning that provides high accuracy, robustness to varying appearances and subtle differences between different yoga poses, and speed sufficient for real-time deployment in practical applications.

Relevant Papers

[1]. Verma, M., Kumawat, S., Nakashima, Y., & Raman, S. (2020). Yoga-82: a new dataset for fine-grained classification of human poses. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 1038-1039).

[2]. Garg, S., Saxena, A., & Gupta, R. (2023). Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application. Journal of Ambient Intelligence and Humanized Computing, 14(12), 16551-16562. https://link.springer.com/article/10.1007/s12652-022-03910-0

[2]. Dittakavi, B., Bavikadi, D., Desai, S. V., Chakraborty, S., Reddy, N., Balasubramanian, V. N., ... & Sharma, A. (2022). Pose tutor: an explainable system for pose correction in the wild. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3540-3549).

[4] Yadav, S. K., Agarwal, A., Kumar, A., Tiwari, K., Pandey, H. M., & Akbar, S. A. (2022). YogNet: A two-stream network for realtime multiperson yoga action recognition and posture correction. *Knowledge-Based Systems*, *250*, 109097.
https://www.sciencedirect.com/science/article/pii/S095070512200541X

Project ID: 015

Project Title: Fine-tune Multi-modal LLaVA Vision and Language Models

Area of Research: Multimodal AI, Large Vision and Language Models

Problem Statement:
With the release of GPT-4o and Gemini, multimodal AI has become an emerging field. A vision-and-language multimodal model, such as LLaVA (Large Language and Vision Assistant), integrates visual information and textual data to perform multimodal tasks that require understanding both images and text. The primary objective of multimodal task is to develop a model that can accurately answer questions about the content of images. This capability has wide-ranging applications, including robots, medical diagnostic system, etc, which allow to attach an image and ask a question about it. With the power of large language models, large vision-and-language models have also achieved significant results. However, there remains a critical need to fine-tune them for specific domains to enhance their accuracy. Fine-tuning allows these models to adapt to domain-specific datasets, such as fine-tuning a model on a medical image dataset to enhance its ability to answer questions related to medical diagnoses. This project aims to fine-tune a large vision-and-language model, LLaVA, on a customized dataset to enhance its ability to answer questions and improve performance.

Dataset:

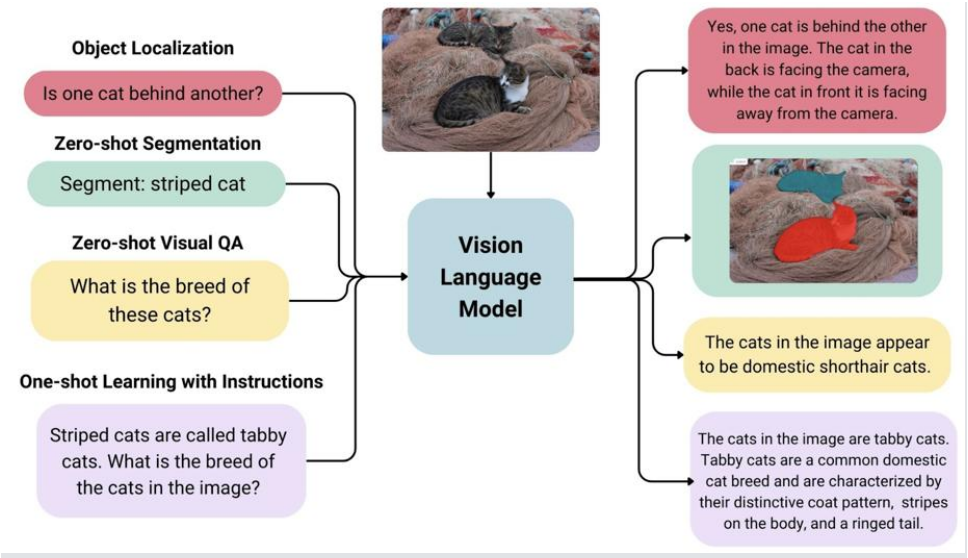


Figure 1. Vision-and-Language Model

The LLaVA-instruct-mix dataset and chess pieces dataset are two multimodal datasets for vision-and-language model fine-tuning and testing. The LLaVA-instruct-mix dataset is a multimodal dataset containing text and image pairs, as well as human-written instructions for vision dialogue tasks. The chess-pieces dataset is a small dataset containing images of chess pieces and related information. It allows to fine-tune the model and test the results to improve the recognition of chess pieces.

Dataset URL:
llava-instruct-mix Dataset: https://huggingface.co/datasets/HuggingFaceH4/llava-instruct-mix-vsft?row=93

Chess_pieces: https://huggingface.co/datasets/Trelis/chess_pieces

Paper: https://arxiv.org/pdf/2304.08485

Task: The primary task of this project is to fine-tune the multimodal LLaVA vision and language models using domain-specific datasets to excel in understanding and responding to domain-specific queries and improving performance.

Relevant Papers
[1]. Liu, H., Li, C., Wu, Q. and Lee, Y.J., 2024. Visual instruction tuning. *Advances in neural information processing systems*, *36*.
https://arxiv.org/pdf/2304.08485

Project ID: 016

Project Title: NLP Classification of FOMC Texts for Market Analysis

Area of Research: Natural Language Processing (NLP)

Problem Statement:
In modern financial markets, the Federal Reserve's (Fed) monetary policy announcements, especially the Federal Open Market Committee (FOMC) statements, significantly influence market trends. Consequently, there is an urgent need for a model capable of extracting policy stances from FOMC texts to better understand and predict their impact on financial markets. This project aims to leverage advanced artificial intelligence techniques to classify sentences in FOMC texts as hawkish, dovish, or neutral. This provides a new tool for financial market analysis, offering market participants a more accurate means of assessing policy stances and aiding in the better understanding and prediction of the FOMC statements' effects on financial markets.

Dataset:

	A	B	C	D
1	index	sentence	year	label
2	487	Instead,	1998	2
3	422	Inflatior	2004	0
4	150	It isn’ t	2021	2
5	110	At some p	2022	1
6	548	Moreover,	2002	0
7	122	Based on	2007	0
8	282	We have l	2020	2

Figure 1. This figure is a partial display of the FOMC text dataset, which spans from 1996 to 2022. The dataset includes text records from FOMC speeches, meeting minutes, and press conferences.

Regarding the meeting minutes, there are 214 documents containing 20,618 target sentences. The press conferences comprise a total of 63 documents with 5,086 target sentences. For speeches, there are 1,026 documents (filtered down to 201), with 12,465 target sentences. The specific links are as follows:

Dataset URL: https://github.com/gtfintechlab/fomc-hawkish-dovish/tree/main/training_data/test-and-training

Dataset paper: https://arxiv.org/pdf/2305.07972

Task: This project proposes the task of classifying Federal Open Market Committee (FOMC) texts. By implementing a "hawkish-dovish" classification task, it allows for a more accurate extraction of monetary policy stances from FOMC texts. This detailed classification aids in understanding and predicting the specific impacts of FOMC statements on financial markets, thereby providing market participants and policymakers with more accurate and valuable analytical tools.

Relevant Papers

[1]. Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10 - Ks[J]. The Journal of finance, 2011, 66(1): 35-65. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2010.01625.x

[2]. Hansen S, McMahon M. Shocking language: Understanding the macroeconomic effects of central bank communication[J]. Journal of International Economics, 2016, 99: S114-S133. https://www.sciencedirect.com/science/article/pii/S0022199615001828

[3]. Araci D. Finbert: Financial sentiment analysis with pre-trained language models[J]. arXiv preprint arXiv:1908.10063, 2019. https://arxiv.org/pdf/1908.10063

[4]. Mathur P, Neerkaje A, Chhibber M, et al. Monopoly: Financial prediction from monetary policy conference videos using multimodal cues[C]//Proceedings of the 30th ACM International Conference on Multimedia. 2022: 2276-2285. https://dl.acm.org/doi/abs/10.1145/3503161.3548380

[5]. Shah A, Paturi S, Chava S. Trillion dollar words: A new financial dataset, task & market analysis[J]. arXiv preprint arXiv:2305.07972, 2023. https://arxiv.org/pdf/2305.07972

Project ID: 017

Project Title: Image Generation

Area of Research: Image Generation, Generative AI, Computer Vision

Problem Statement: Generative AI has made significant strides in creating realistic images, with diffusion models emerging as a powerful approach in this domain. Diffusion models, which gradually transform a simple noise distribution into complex data distributions, have demonstrated remarkable capabilities in generating high-quality images. This project aims to build a diffusion model from scratch, specifically tailored to generate realistic images of cars. By training the model on the Stanford Cars Dataset, we aim to produce a generative image model that can create new car images, providing deeper understanding of the mechanics of diffusion models and contributing to advancements in generative AI.

Dataset:

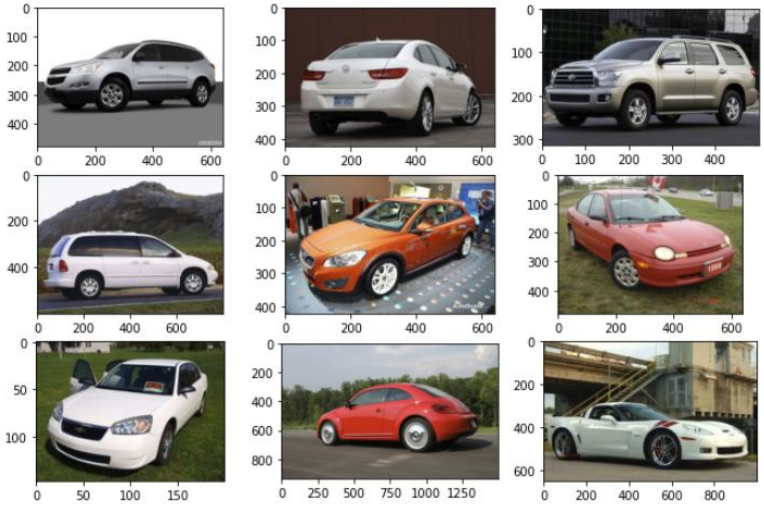


Figure 1. StandordCars Dataset

StandordCars Dataset contains 16,185 images of 196 classes of cars. The images are divided almost 50-50 between training and testing, with 8,144 training images and 8,041 testing images. Categories are typically at the make, model, and year level.

Dataset URL: https://www.kaggle.com/datasets/jessicali9530/stanford-cars-dataset

Paper: https://arxiv.org/pdf/2006.11239

Task: The primary task of this project is to develop a generative image model using a diffusion approach and train the model on SatndordCars Dataset to generate new car images.

Relevant Papers

[1]. Ho, J., Jain, A. and Abbeel, P., 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, pp.6840-685.

https://arxiv.org/pdf/2006.11239

[2]. Dhariwal, P. and Nichol, A., 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, *34*, pp.8780-8794.

https://arxiv.org/abs/2105.05233

Project ID: 018

Project Title: Kidney (and Kidney tumour) segmentation

Area of Research: Computer Vision, Medical Image analysis

Problem Statement: Accurate segmentation of kidneys and kidney tumors in computed tomography (CT) scans is critical for diagnosis, surgical planning, and treatment monitoring in patients with renal tumors. Manual segmentation of these structures is time-consuming, prone to variability among radiologists, and can be challenging due to the complexity of abdominal anatomy and the variability in tumor appearance and location.

Dataset:

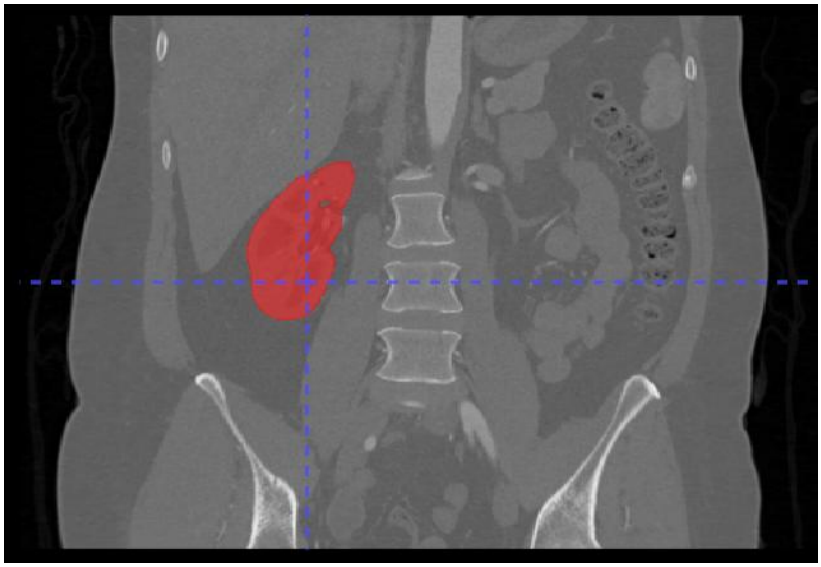


Figure 1Kidney Segmentation Sample

Dataset URL: https://kits19.grand-challenge.org/data

Dataset paper: Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M. and Dean, J., 2019. The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. *arXiv preprint arXiv:1904.00445*.

Task: Semantic Segmentation of Kidney, Medical Image Segmentation. Please note that 3D semantic segmentation and kidney tumour segmentation are not mandatory. All data (Image and ground truth) could be converted into 2D slices and unlabeled data/slices could be ignored.

Relevant Papers

1. Taha, A., Lo, P., Li, J. and Zhao, T., 2018. Kid-net: convolution networks for kidney vessels segmentation from ct-volumes. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11 (pp. 463-471). Springer International Publishing.

2. Sathianathen, N.J., Heller, N., Tejpaul, R., Stai, B., Kalapara, A., Rickman, J., Dean, J., Oestreich, M., Blake, P., Kaluzniak, H. and Raza, S., 2022. Automatic segmentation of kidneys and kidney tumors: The KiTS19 international challenge. Frontiers in Digital Health, 3, p.797607.

3. Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

Project ID: 019

Project Title: Visual Question Answering

Area of Research: Multimodal AI

Problem Statement:
With the release of GPT-4o and Gemini, multimodal AI has become an emerging field. One of the fundamental research tasks is Visual Question Answering (VQA), a rapidly evolving field that combines computer vision and natural language processing. The primary objective of VQA is to develop a model that can accurately answer questions about the content of images. This capability has wide-ranging applications, including robots and systems like GPT-4o, Gemini, etc, which allow you to attach an image and ask a question about it. Despite significant advancements in computer vision and natural language processing, achieving multimodal tasks such as VQA remains a challenging task due to the need for the system to understand both the visual context of the image and the semantic meaning of the question. The goal of this project is to develop a VQA model that can interpret images and answer related questions with high accuracy.

Dataset:

Figure 1. VQAv2 Dataset. Ask a question about an image and require the model to provide a correct answer.

The VQAv2 dataset [1], and GQA[2] are two most popular benchmark datasets for visual question answering. The details about the dataset can be found in [1][2] and be downloaded from the URL given below (you only need to choose one dataset):

Dataset URL:
VQAv2: https://visualqa.org/ or https://huggingface.co/datasets/HuggingFaceM4/VQAv2
GQA: https://cs.stanford.edu/people/dorarad/gqa/download.html

Dataset paper: https://arxiv.org/pdf/1612.00837 and https://arxiv.org/pdf/1902.09506

Task: To develop a robust VQA system using deep neural networks that can accurately answer a wide range of questions based on images. The system shall be able to answer different types of VQA questions, such as open-ended questions, multiple-choice questions, counting questions, and yes/no questions.

Relevant Papers
[1]. Goyal, Y., Khot, T., Summers-Stay, D., Batra, D. and Parikh, D., 2017. Making the v in vqa matter: Elevating the role of image understanding in visual question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6904-6913).
https://arxiv.org/pdf/1612.00837
[2]. Hudson, D.A. and Manning, C.D., 2019. Gqa: A new dataset for real-world visual reasoning and compositional question answering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6700-6709). https://arxiv.org/pdf/1902.09506
[3] Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C.L. and Parikh, D., 2015. Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision* (pp. 2425-2433).
https://arxiv.org/pdf/1505.00468
[4] Zhang, P., Goyal, Y., Summers-Stay, D., Batra, D. and Parikh, D., 2016. Yin and yang: Balancing and answering binary visual questions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5014-5022). https://arxiv.org/pdf/1511.05099

Project ID: 020

Project Title: Image Captioning

Area of Research: Image Captioning, Computer Vision, Natural Language Processing

Problem Statement: Image captioning is fundamental task for recent cutting-edge multimodal AI and image generation applications. It serves as a bridge by linking visual content to natural language, enhancing the interaction between images and text. The goal of this project is to develop a model that can generate accurate textual descriptions for given images. Despite significant advancements in both computer vision and natural language processing, existing models often struggle with providing detailed and accurate captions, especially when dealing with diverse and complex images. Fine-tuning image captioning models on specific datasets can significantly improve their performance by enabling them to generate more precise and contextually appropriate descriptions. This process ensures the models can better understand the visual content and translate it into meaningful textual representations.

Dataset:

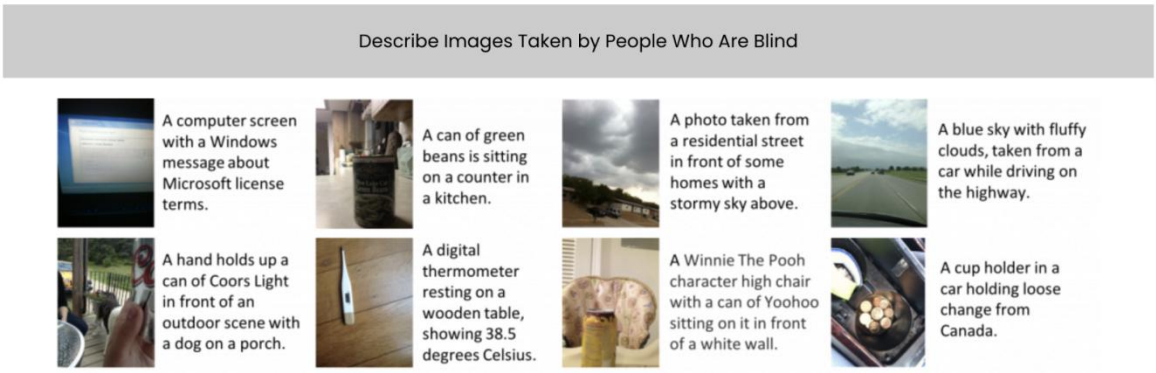


Figure 1. VizWiz-Captions

VizWiz-Captions is an image captioning dataset specifically designed to represent the real use case of people who are blind relying on image captioning services. This dataset consists of 39,181 images, each originating from individuals who are blind and paired with 5 corresponding captions. The primary task addressed by VizWiz-Captions is predicting a suitable caption given an image. This dataset aims to educate more people about the technological needs of blind people while providing a new opportunity for researchers to develop assistive technologies that eliminate accessibility barriers for blind people.

Dataset URL: https://vizwiz.org/tasks-and-datasets/image-captioning/

Paper: https://arxiv.org/abs/2002.08565

Task: The primary task of this project is to build an image captioning model using domain-specific datasets or fine-tune the pre-trained image captioning model using the curated dataset. The project involves training the model to better understand and generate captions that are specific to the images in the dataset.

Relevant Papers

[1]. Gurari, D., Zhao, Y., Zhang, M. and Bhattacharya, N., 2020. Captioning images taken by people who are blind. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16* (pp. 417-434). Springer International Publishing.
https://arxiv.org/pdf/2002.08565

[2]. Simons, R.N., Gurari, D. and Fleischmann, K.R., 2020. " I Hope This Is Helpful" Understanding Crowdworkers' Challenges and Motivations for an Image Description Task. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), pp.1-26.
https://dl.acm.org/doi/pdf/10.1145/3415176

Project ID: 021

Project Title: Emotion Classification using Tweets.

Area of Research: Natural Language Processing.

Problem Statement: Emotions are expressed in nuanced ways, which varies by collective or individual experiences, knowledge, and beliefs. Therefore, to understand emotion, as conveyed through text, a robust mechanism capable of capturing and modelling different linguistic nuances and phenomena is needed. Hence, the task of this project is to develop deep learning based solutions that can accurately identify emotion presented in textual data.

Dataset:

Emotions	Amount	Hashtags
sadness	214,454	#depressed, #grief
joy	167,027	#fun, #joy
fear	102,460	#fear, #worried
anger	102,289	#mad, #pissed
surprise	46,101	#strange, #surprise
trust	19,222	#hope, #secure
disgust	8,934	#awful, #eww
anticipation	3,975	#pumped, #ready

Figure 1. Data statistics of EMOTION dataset

EMOTION[1] is a dataset of English Twitter messages with six basic emotions: anger, fear, joy, love, sadness, and surprise.

Dataset URL: https://huggingface.co/datasets/dair-ai/emotion

Dataset paper: https://aclanthology.org/D18-1404/

Task: The goal of this project is to perform text classification using advanced artificial intelligence algorithms that yield increased performance on relevant metrics. The text classification for this project is correctly labelling a tweet for emotion expressed by the author.

Relevant Papers

[1]. Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. 2018. CARER: Contextualized Affect Representations for Emotion Recognition. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3687–3697, Brussels, Belgium. Association for Computational Linguistics. https://aclanthology.org/D18-1404/

[2]. Muhammad Abdul-Mageed and Lyle Ungar. 2017. EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 718–728, Vancouver, Canada. Association for Computational Linguistics. https://aclanthology.org/P17-1067/

[3]. Federico Bianchi, Debora Nozza, and Dirk Hovy. 2021. FEEL-IT: Emotion and Sentiment Classification for the Italian Language. In Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 76–83, Online. Association for Computational Linguistics. https://aclanthology.org/2021.wassa-1.8/

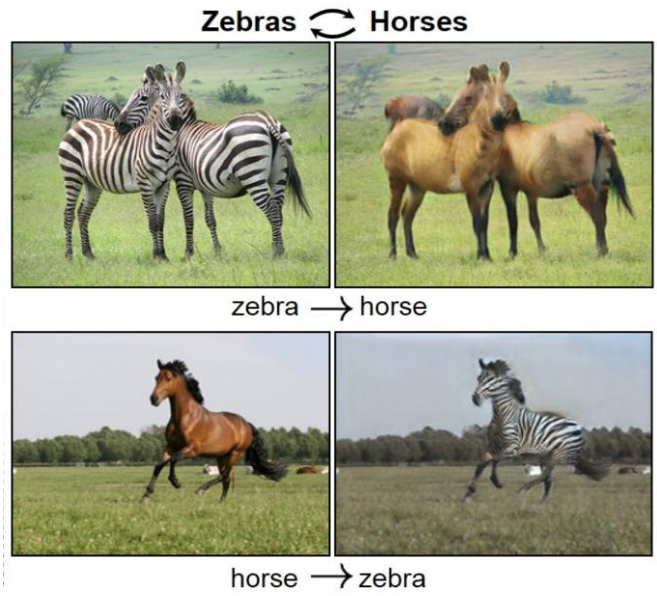
Project ID: 022

Project Title: Style Transfer: Transforming Horses into Zebras

Area of Research: Generative AI

Problem Statement: Style transfer is a fascinating technique that allows for the transformation of an image's style while preserving its essential content. Rapid and accurate application of style transfer is essential for various creative and practical applications, such as generating art, enhancing images, and transforming animal appearances. One intriguing application of this technique is converting images of horses into images of zebras. However, achieving realistic transformations while maintaining the structural integrity of the original image can be challenging. The goal of this project is to explore and implement an advanced style transfer algorithm that can accurately transform horse images into zebra images, providing high-quality, realistic results.

Dataset:



[换行符]

Figure 1. Example images if the zebra2horse and horse2zebra task, with dataset sample for zebra and horse.

Horse2zebra dataset [1] is a benchmark dataset for style transfer. The details about the dataset can be found in [1] and be downloaded from the URL given below:
Dataset URL: https://www.tensorflow.org/datasets/catalog/cycle_gan

Task: To develop an automatic horse-to-zebra transformation system using neural networks and deep learning that provides high accuracy, maintains the structural integrity of the original image, and produces realistic transformations. The system should be robust to varying appearances of horses and zebras and efficient enough to look as real as possible.

Relevant Papers

[1]. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232). https://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.html
[2]. Tang, H., Liu, H., Xu, D., Torr, P. H., & Sebe, N. (2021). Attentiongan: Unpaired image-to-image translation using attention-guided generative adversarial networks. IEEE transactions on neural networks and learning systems, 34(4), 1972-1987.
https://ieeexplore.ieee.org/abstract/document/9527389

[3]. Kim, T., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017, July). Learning to discover cross-domain relations with generative adversarial networks. In *International conference on machine learning* (pp. 1857-1865). PMLR.
https://proceedings.mlr.press/v70/kim17a

Project ID: 023

Project Title: Financial sentiment classification.

Area of Research: NLP

Problem Statement: The public sentiment or outlook of a stock is a powerful indicator of future prices. These could be tweets, Reddit posts, or news headlines among others are valuable sources of signals to predict stock movements. However, the large volume of data makes manually discerning if public sentiment for stock from news headlines or social media posts is positive or negative infeasible. Thus, the use of natural language processing in finance has grown as financial institutions implement sentiment analysis to gauge public opinion on stocks. Developing an automated sentiment analysis model would allow for the processing of large volumes of news articles in real time, which is crucial for financial markets as time sensitivity is important for trading decisions. The goal of this project is to create a model that can accurately classify the sentiment or a news headline, such that financial professionals can gain immediate insights into market sentiment, leading to better investment strategies and risk management.

Dataset:

positive	HELSINKI (AFX) - Shares closed higher , led by Nokia after it announced plans to team up with Sany...
positive	Incap Contract Manufacturing Services Pvt Ltd , a subsidiary of Incap Corporation of Finland , plans...

Figure 1. Two samples of data from the financial phrase bank dataset.

Dataset URL: https://huggingface.co/datasets/takala/financial_phrasebank

Dataset paper: https://arxiv.org/abs/1307.5336

Task: To build a deep-learning model that can best classify financial news headlines into positive, neutral, or negative sentiment classes. The model should be accurate and generalizable to different sources of data.

Relevant Papers

[1. Pekka Malo, Ankur Sinha, Pyry Takala, Pekka Korhonen, Jyrki Wallenius: “Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts”, 2013; [http://arxiv.org/abs/1307.5336 arXiv:1307.5336].

[2]. Md Parvez Mollah: “An LSTM model for Twitter Sentiment Analysis”, 2022; [http://arxiv.org/abs/2212.01791 arXiv:2212.01791].

[3]. Dogu Araci: “FinBERT: Financial Sentiment Analysis with Pre-trained Language Models”, 2019; [http://arxiv.org/abs/1908.10063 arXiv:1908.10063].

[4]. Jean Lee, Hoyoul Luis Youn, Josiah Poon, Soyeon Caren Han: “StockEmotions: Discover Investor Emotions for Financial Sentiment Analysis and Multivariate Time Series”, 2023; [http://arxiv.org/abs/2301.09279 arXiv:2301.09279].

Project ID: 024

Project Title: Predicting movie genres from synopsis.

Area of Research: Natural language processing (NLP)

Problem Statement: Being able to sort movies by genre or specific tags is a valuable tool to help enhance user experience on streaming services by providing better recommendations and simplifying the navigation of extensive film libraries. The rise of online movie streaming has presented a challenge in organising very large databases of movies. The goal of this project is to create a text classification deep-learning model that can accurately classify movie synopses into relevant tags or genres. Developing an automated system is advantageous over manual classification as it can be easily scaled to huge sets of data. Moreover, as more movies are added to the database, the model can maintain up-to-date genre categorizations without the need for constant human intervention. This enhances the user experience by providing accurate and timely recommendations.

Dataset:

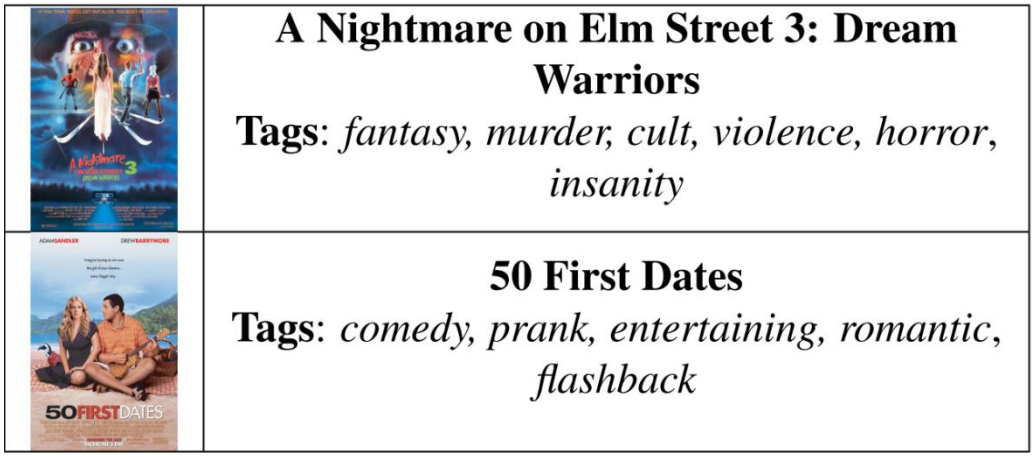


Figure 1. An example of tag assignments to movies from the MPST dataset.

Dataset URL: https://ritual.uh.edu/mpst-2018/

Dataset paper: https://aclanthology.org/L18-1274

Task: To develop a neural network and deep learning model to classify movie synopsis text into relevant genre categories. The model should be accurate and robust to different synopsis lengths and writing styles.

Relevant Papers

[1]. Quan Hoang: “Predicting Movie Genres Based on Plot Summaries”, 2018; [http://arxiv.org/abs/1801.04813 arXiv:1801.04813].

[2]. Ertugrul, Ali Mert & KARAGOZ, Pinar. (2018). Movie Genre Classification from Plot Summaries Using Bidirectional LSTM. 10.1109/ICSC.2018.00043. https://www.researchgate.net/publication/322929271_Movie_Genre_Classification_from_Plot_Summaries_Using_Bidirectional_LSTM

[3]. Wang, Jingcheng. (2020). Using Machine Learning to Identify Movie Genres through Online Movie Synopses. 1-6. 10.1109/ITCA52113.2020.00008. https://www.researchgate.net/publication/351418872_Using_Machine_Learning_to_Identify_Movie_Genres_through_Online_Movie_Synopses

Project ID: 025

Project Title: Lung Nodule Detection Using Deep Learning

Area of Research: Computer Vision (CV)

Problem Statement:

The project aims to develop a deep learning model for accurately detecting chest nodules in chest radiographs. Chest nodule detection is a critical task in radiology, as early and precise identification of nodules can significantly improve patient outcomes by facilitating timely interventions.

Dataset:

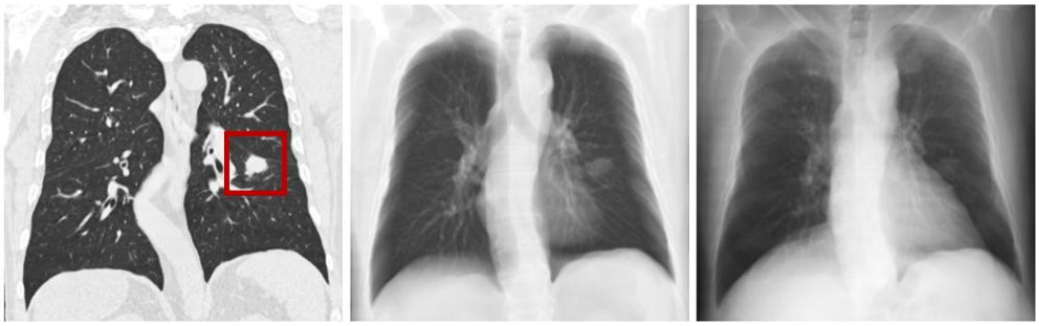


Figure 1: example images of Node21 dataset.

NODE21 public CXR training dataset consists of frontal chest radiographs with annotated bounding boxes around nodules. It consists of 4882 frontal chest radiographs where 1134 CXR images (1476 nodules) are annotated with bounding boxes around nodules and the remaining 3748 images are free of nodules hence represent the negative class.

Dataset URL: https://node21.grand-challenge.org/

https://zenodo.org/records/5548363

Task:

The detection task in the NODE21 dataset involves identifying and localizing nodules within chest radiograph using bounding boxes. This approach provides precise coordinates that outline the location and size of each nodule, which is crucial for early diagnosis and treatment planning in lung disease.

Relevant Papers:

1. Behrendt, Finn, et al. "A systematic approach to deep learning-based nodule detection in chest radiographs." Scientific reports 13.1 (2023): 10120.

2. Sogancioglu, Ecem, et al. "Nodule detection and generation on chest X-rays: NODE21 Challenge." arXiv preprint arXiv:2401.02192 (2024).

3. Behrendt, Finn, et al. "Nodule Detection in Chest Radiographs with Unsupervised Pre-Trained Detection Transformers." 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). IEEE, 2023.

Project ID: 026

Project Title: Multi-Class Hate Speech and Offensive Language Detection in Social Media

Area of Research: Natural Language Processing (NLP)

Problem Statement: Hate speech and offensive language on social media platforms can lead to online harassment, discrimination, and a toxic online environment. Automatically identifying and categorizing such content is crucial for effective content moderation. However, distinguishing between hate speech, offensive language, and normal speech can be challenging, as the boundaries between these categories are often blurred. The goal of this project is to develop a deep learning-based system that can accurately classify social media posts into three categories: hate speech, offensive language, and normal speech. This will enable social media platforms to better understand the nature of potentially problematic content and take appropriate actions.

Dataset:

{"1179055004553900032_twitter": {"post_id": "1179055004553900032_twitter", "annotators": [{"label": "normal", "annotator_id": 1, "target": ["None"]}, {"label": "normal", "annotator_id": 2, "target": ["None"]}, "1179088797964763136_twitter": {"post_id": "1179088797964763136_twitter", "annotators": [{"label": "hatespeech", "annotator_id": 1, "target": ["Asian"]}, {"label": "offensive", "annotator_id": 4, "target": ["Asian"]}, {"label": "hatespeech", "annotator_id": 3, "target": ["Asian"]}], "rationales": [[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0], [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0], [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]], "post_tokens": ["<user>", "i", "am", "bit", "confused", "coz", "chinese", "ppl", "can", "not", "access", "twitter", "thn", "how", "this", "ching", "chong", "using", "it", "i", "think", "he", "pakistani", "\ud83e\udd14", "\ud83e\udd14", "\ud83e\udd14"]},

Figure. Examples of annotated text data for sentiment analysis from social media.

(Each entry includes a post ID, the text of the post, annotators' IDs, targets, labels (normal, offensive, hate speech), rationales, and post tokens).

Dataset URL: https://github.com/hate-alert/HateXplain/tree/master/Data

Task: Develop a deep learning model for multi-class classification of social media posts into three categories: hate speech, offensive language, and normal speech.

Relevant Papers

[1] Mathew, B., Saha, P., Yimam, S. M., Biemann, C., Goyal, P., & Mukherjee, A. (2021). HateXplain: A benchmark dataset for explainable hate speech detection. Proceedings of the AAAI Conference on Artificial Intelligence, 35(17), 14867-14875. https://doi.org/10.1609/aaai.v35i17.17745

[2] Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 512-515. https://ojs.aaai.org/index.php/ICWSM/article/view/14955

[3] Zampieri, M., Nakov, P., Rosenthal, S., Atanasova, P., Karadzhov, G., Mubarak, H., Derczynski, L., Pitenis, Z., & Çöltekin, Ç. (2020). SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020). Proceedings of the Fourteenth Workshop on Semantic Evaluation, 1425-1447. https://aclanthology.org/2020.semeval-1.188/

[4] Mozafari, M., Farahbakhsh, R., & Crespi, N. (2020). A BERT-based transfer learning approach for hate speech detection in online social media. In H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, & L. M. Rocha (Eds.), Complex Networks and Their Applications VIII (pp. 928-940). Springer. https://doi.org/10.1007/978-3-030-36687-2_77

Project ID: 027

Project Title: Fashion Item Classification using Deep Learning

Area of Research: Computer Vision

Problem Statement: Automated classification of fashion items is crucial for various applications in the fashion industry, such as product categorization, recommendations, and search optimization. However, accurately classifying fashion items can be challenging due to the wide variety of styles, designs, and visual similarities between different categories. This project aims to address this challenge by leveraging deep learning techniques, particularly convolutional neural networks (CNNs), to develop a robust and efficient system for classifying fashion items based on images. By training a CNN model on the UT Fashion 100 dataset, which consists of a diverse collection of fashion item images, the goal is to create a system that can accurately identify different categories of fashion items, facilitating automated processes in the fashion industry and enhancing the user experience in fashion-related applications.

Dataset: **UT Fashion 100 dataset**

Figure. Example images from the UT Fashion 100 dataset showcasing different fashion item categories.

(Each image in the dataset is labelled with the corresponding fashion item category in the json file.)

Task:

a. Develop a deep learning-based image classification model using CNNs to accurately classify fashion items into four different categories – shoes, clothing, accessories, bags from input images.

(OR)

b. Develop a deep learning-based image classification model using CNNs to accurately classify fashion items into four different categories – shoes, clothing, accessories, bags and find their sub categories from the main categories.

Relevant Papers:

[1]. Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). DeepFashion: Powering robust clothes recognition and retrieval with rich annotations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1096-1104). https://doi.org/10.1109/CVPR.2016.124

[2]. Corbiere, C., Ben-Younes, H., Rame, A., & Ollion, C. (2017). Leveraging weakly annotated data for fashion image retrieval and label prediction. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 2268-2274). https://doi.org/10.1109/ICCVW.2017.266

[3]. Zou, X., Kong, X., Wong, W., Wang, C., Liu, Y., & Cao, Y. (2019). FashionAI: A hierarchical dataset for fashion understanding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0). https://doi.org/10.1109/CVPRW.2019.00078

Project ID: 028

Project Title: Environmental Sound Classification

Area of Research: Computer Vision (Audio Signal Processing)

Problem Statement: Environmental sound classification is a challenging task in the field of audio signal processing and machine learning. It has various applications, such as acoustic monitoring, surveillance systems, and smart home assistants. However, accurately classifying environmental sounds can be difficult due to the wide variety of sound events, background noise, and acoustic variations. This project aims to address this challenge by leveraging deep learning techniques, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to develop a robust and efficient system for classifying environmental sounds based on audio features. By training a deep learning model on the ESC-50 dataset, the goal is to create a system that can accurately identify different types of environmental sounds, enabling better understanding and interpretation of acoustic scenes.

Dataset: The dataset consists of 5-second-long recordings organized into 50 semantical classes (with 40 examples per class) loosely arranged into 5 major categories

Dataset URL: https://github.com/karolpiczak/ESC-50

Task: Develop a deep learning-based environmental sound classification model using CNNs and RNNs to accurately classify audio clips into their respective sound categories.

Relevant Papers:

[1]. Piczak, K. J. (2015). ESC: Dataset for environmental sound classification. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 1015-1018). http://karol.piczak.com/papers/Piczak2015-ESC-Dataset.pdf

[2]. Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Processing Letters, 24(3), 279-283. https://doi.org/10.1109/LSP.2017.2657381

[3]. Zhang, Z., Xu, S., Zhang, S., Qiao, T., & Cao, S. (2019). Learning attentive representations for environmental sound classification. IEEE Access, 7, 130327-130339. https://doi.org/10.1109/ACCESS.2019.2939495

[4]. Boddapati, V., Petef, A., Rasmusson, J., & Lundberg, L. (2017). Classifying environmental sounds using image recognition networks. Procedia Computer Science, 112, 2048-2056. https://doi.org/10.1016/j.procs.2017.08.250

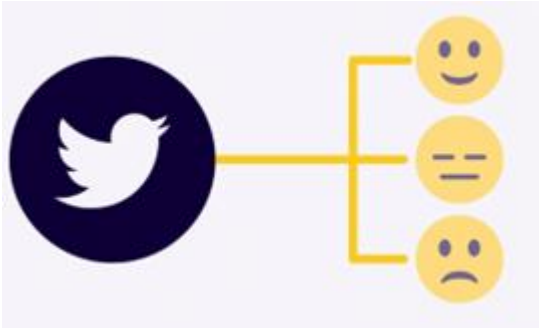
Project ID: 029

Project Title: Sentiment Analysis towards COVID-19 on Twitter

Area of Research: Natural Language Processing (NLP).

Problem Statement: Sentiment analysis of Tweets is one of the most popular tasks in the NLP space; it has many practical applications including stock prediction, box office results estimation, survey of popular opinions and content recommendation. This project involves estimating the sentiment of tweets about the COVID-19 pandemic to track people's opinions on major events surrounding it.

Dataset:



CovidSENTI is a large-scale benchmark dataset for COVID-19 Sentiment Analysis. It contains a balanced distribution of positive, neutral and negative tweets about the pandemic, and can be downloaded at the link below.

Dataset URL: https://github.com/usmaann/COVIDSenti
Dataset Paper*: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9340540*

Task: Classify the sentiment of tweets involving the COVID-19 pandemic into three classes: Positive, Negative and Neutral.

Relevant Papers

[1].Naseem, Usman, et al. "COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis." IEEE transactions on computational social systems 8.4 (2021): 1003-1015. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9340540

[3]. Joloudari, Javad Hassannataj, et al. "BERT-deep CNN: State of the art for sentiment analysis of COVID-19 tweets." Social Network Analysis and Mining 13.1 (2023): 99. https://link.springer.com/article/10.1007/s13278-023-01102-y

[3]. Braig, Niklas, et al. "Machine learning techniques for sentiment analysis of COVID-19-related twitter data." IEEE Access 11 (2023): 14778-14803. https://ieeexplore.ieee.org/iel7/6287639/6514899/10035946.pdf

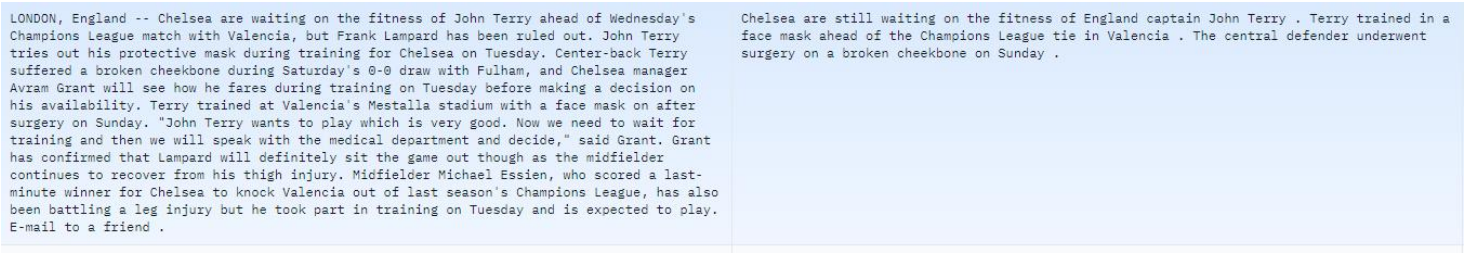
Project ID: 030

Project Title: Summarising news articles.

Area of Research: Natural language processing (NLP)

Problem Statement: In recent years, the volume of news generated on the internet has become overwhelming for readers to stay informed. As readers face time constraints, summaries of news articles to quickly explain the key ideas are essential in allowing readers to be more time efficient. Additionally, the large number of news media that is available online has meant that news companies are competing to be able to keep their readers up to date efficiently. Unfortunately, not all news sources come with summaries or dot points that easily convey ideas immediately. The goal of this project is to develop an automated text summarization model for news articles that can take in a longer piece of text and generate a summary of the key ideas from that text while maintaining the correct context of the information and being coherent to the reader. Such models can be used by news aggregator sites that take in huge quantities of news articles daily where manually writing summaries is infeasible. A text summarization model would be vital in improving reader experience by increasing time efficiency.

Dataset:



LONDON, England -- Chelsea are waiting on the fitness of John Terry ahead of Wednesday's Champions League match with Valencia, but Frank Lampard has been ruled out. John Terry tries out his protective mask during training for Chelsea on Tuesday. Center-back Terry suffered a broken cheekbone during Saturday's 0-0 draw with Fulham, and Chelsea manager Avram Grant will see how he fares during training on Tuesday before making a decision on his availability. Terry trained at Valencia's Mestalla stadium with a face mask on after surgery on Sunday. "John Terry wants to play which is very good. Now we need to wait for training and then we will speak with the medical department and decide," said Grant. Grant has confirmed that Lampard will definitely sit the game out though as the midfielder continues to recover from his thigh injury. Midfielder Michael Essien, who scored a last-minute winner for Chelsea to knock Valencia out of last season's Champions League, has also been battling a leg injury but he took part in training on Tuesday and is expected to play. E-mail to a friend .	Chelsea are still waiting on the fitness of England captain John Terry . Terry trained in a face mask ahead of the Champions League tie in Valencia . The central defender underwent surgery on a broken cheekbone on Sunday .

Figure 1. An example datapoint from the cnn_dailymail dataset. On the left is the full article and on the right side is the human-written summary.

Dataset URL: https://huggingface.co/datasets/ccdv/cnn_dailymail

Dataset paper: https://arxiv.org/abs/1602.06023v5

Task: To create a deep-learning model for extractive or abstractive text summarisation on news articles. The model should aim to accurately convey the key ideas of a text concisely while maintaining coherence and being robust to different writing styles.

Relevant Papers

[1]. Danqi Chen, Jason Bolton, Christopher D. Manning: “A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task”, 2016; [http://arxiv.org/abs/1606.02858 arXiv:1606.02858].

[2]. Yang Liu: “Fine-tune BERT for Extractive Summarization”, 2019; [http://arxiv.org/abs/1903.10318 arXiv:1903.10318].

[3]. Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer: “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension”, 2019; [http://arxiv.org/abs/1910.13461 arXiv:1910.13461].

[4]. Tohida Rehman, Suchandan Das, Debarshi Kumar Sanyal, Samiran Chattopadhyay: “Abstractive Text Summarization using Attentive GRU based Encoder-Decoder”, 2023, https://link.springer.com/chapter/10.1007/978-981-19-4831-2_56(2022); [http://arxiv.org/abs/2302.13117 arXiv:2302.13117]. DOI:

Project ID: 031

Project Title: Multiple Instance Captioning for Histopathology Images Using ARCH Dataset

Area of Research: Computer Vision and NLP

Problem Statement: Histopathology is a critical field in medical diagnosis, involving the examination of tissue samples under a microscope to identify diseases. Automated captioning of histopathology images can assist pathologists by providing accurate descriptions of tissue morphology, diagnostic information, and other relevant details. The ARCH dataset provides a unique opportunity to develop models that can generate dense captions for histopathology images, covering various tissue types, stains, and pathologies. This project aims to develop a deep learning model for multiple instance captioning of histopathology images using the ARCH dataset, enhancing the accuracy and efficiency of pathological diagnoses.

Dataset: ARCH Dataset [1]

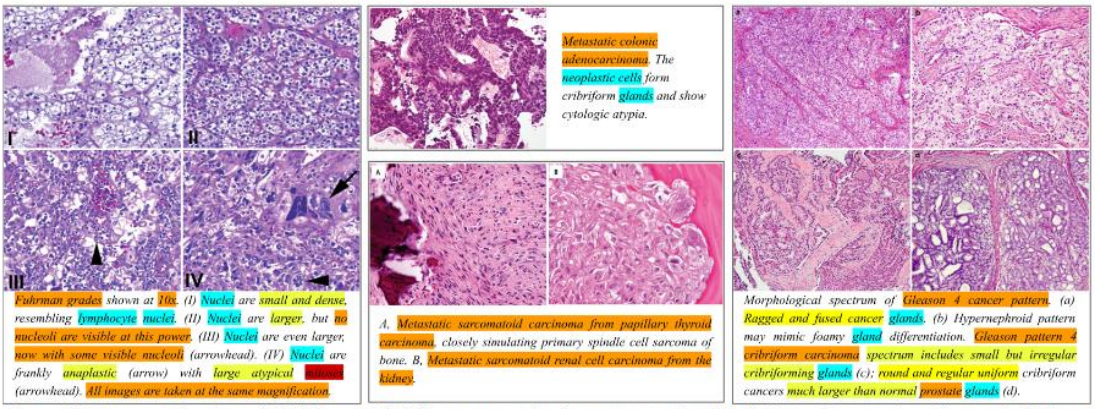


Figure 1: Four samples from ARCH, a multiple instance captioning computational pathology dataset. Samples on the left and right each consist of four image instances with a single caption; top-middle shows an image-caption pair while bottom-middle contains two image instances with a single caption. Labeled in color are examples of common tasks within computational pathology: diagnostic (orange); detection & classification (cyan); descriptive (yellow); special cell detection (red).

The ARCH dataset includes dense captions for histopathology images, covering a range of stains, tissue types, and pathologies. The dataset consists of multiple image instances grouped into bags, each annotated with a single, detailed caption. This structure allows for the exploration of complex relationships between images and their corresponding textual descriptions.

Dataset URL: https://warwick.ac.uk/fac/cross_fac/tia/data/arch

Task:
Develop an automated system for generating captions for histopathology images.

Relevant Papers

[1]. Gamper, J., & Rajpoot, N. M. (2021). "Multiple Instance Captioning: Learning Representations from Histopathology Textbooks and Articles." arXiv preprint arXiv:2103.05121.

Project ID: 032

Project Title: Colour Stain Normalisation of Histopathology Images

Area of Research: Computer Vision (CV)

Problem Statement: Stain color variation in Hematoxylin and Eosin (H&E) stained histopathology images poses significant challenges in effectively training deep learning-based algorithms. These variations can arise from differences in staining protocols, tissue preparation methods, and scanning devices, leading to inconsistent color representation across different images. This inconsistency can adversely affect the performance and generalizability of deep learning models, which rely heavily on uniform input data for accurate training and prediction. To address this issue, various stain normalization methods have been proposed. Among these, generative adversarial networks (GANs) have shown promising results by learning to map the color distribution of source images to a target distribution, thereby reducing the variability introduced by staining differences. Despite the advancements, there remains a need for further exploration and refinement of these methods to ensure robust and reliable normalization across diverse datasets. The public histopathology image dataset CAMELYON17, which includes a wide array of H&E stained whole-slide images from different medical centers, presents an ideal testing ground for developing and validating such normalization techniques. By leveraging this dataset, researchers can systematically evaluate the effectiveness of GAN-based stain normalization methods in mitigating color variability and improving the performance of deep learning models in histopathological image analysis.

Dataset:

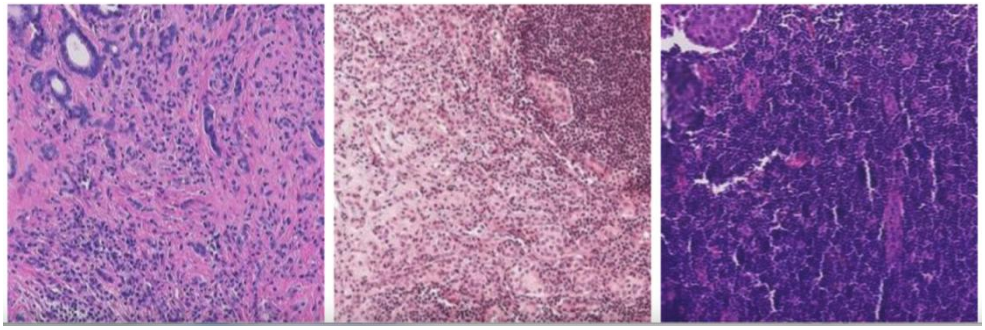


Figure 1. Example images of CAMELYON16 dataset. Colour heterogeneity can be observed within the same dataset.

The CAMELYON16 dataset [1] is a benchmark dataset for histopathology images. The details about the dataset can be found in [1] and can be downloaded from the URL given below:
Dataset URL: https://camelyon17.grand-challenge.org/Data/

Task: To develop an automatic color stain normalization method using generative adversarial networks (GANs) that provides high accuracy and robustness in normalizing the colors of histopathology images. This advancement aims to enhance the reliability and applicability of medical image processing in real-world clinical settings.

Relevant Papers

[1]. Cong, C., Liu, S., Di Ieva, A., Pagnucco, M., Berkovsky, S., & Song, Y. (2022). Colour adaptive generative networks for stain normalisation of histopathology images. Medical Image Analysis, 82, 102580.
https://shlomo-berkovsky.github.io/files/pdf/MedIA22.pdf
[2]. Cong, C., Liu, S., Di Ieva, A., Pagnucco, M., Berkovsky, S., & Song, Y. (2021, April). Texture enhanced generative adversarial network for stain normalisation in histopathology images. In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (pp. 1949-1952). IEEE.
https://shlomo-berkovsky.github.io/files/pdf/ISBI21.pdf
[3]. Cho H, Lim S, Choi G, Min H. Neural Stain-Style Transfer Learning Using Gan for Histopathological Images. arXiv 2017; arXiv: 1710.08543.
https://arxiv.org/pdf/1710.08543
[4]. Shrivastava A, Adorno W, Sharma Y, Ehsan L, Ali SA, Moore SR, et al. Self-Attentive Adversarial Stain Normalization. arXiv Preprint arXiv: 1909.01963; 2019.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8528268/
[5]. Salehi P, Chalechale A. Pix2Pix-based Stain-to-Stain Translation: A Solution for Robust Stain Normalization in Histopathology Images Analysis. International Conference on Machine Vision and Image Processing (MVIP); 2020. p. 1-7.
https://arxiv.org/pdf/2002.00647

Project ID: 033

Project Title: Digital Retinal Images for Vessel Segmentation

Area of Research: Computer Vision (CV)

Problem Statement: The DRIVE database has been established to enable comparative studies on segmentation of blood vessels in retinal images. Retinal vessel segmentation and delineation of morphological attributes of retinal blood vessels, such as length, width, tortuosity, branching patterns and angles are utilized for the diagnosis, screening, treatment, and evaluation of various cardiovascular and ophthalmologic diseases such as diabetes, hypertension, arteriosclerosis and chorodial neovascularization. Automatic detection and analysis of the vasculature can assist in the implementation of screening programs for diabetic retinopathy, can aid research on the relationship between vessel tortuosity and hypertensive retinopathy, vessel diameter measurement in relation with diagnosis of hypertension, and computer-assisted laser surgery. Automatic generation of retinal maps and extraction of branch points have been used for temporal or multimodal image registration and retinal image mosaic synthesis. Moreover, the retinal vascular tree is found to be unique for each individual and can be used for biometric identification.

Dataset:

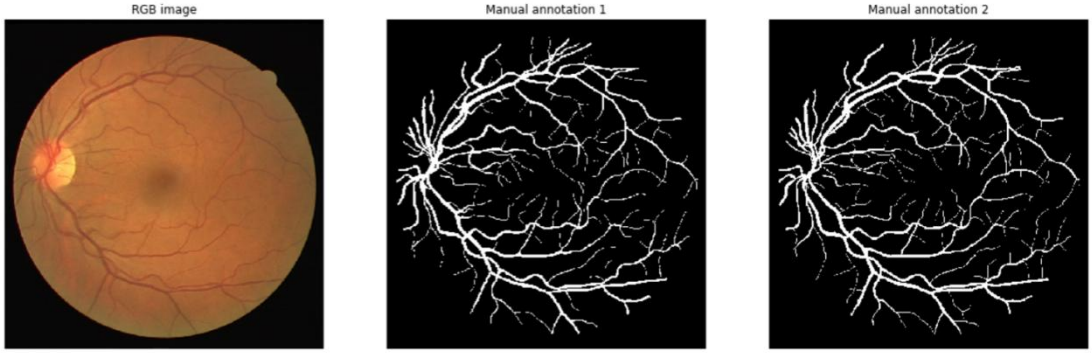


Figure 1. Example images of DRIVE dataset. The first image represents the original retinal image and the other two are annotations by professional experts.

The DRIVE dataset [1] is a benchmark dataset for digital retinal of the vessel. The details about the dataset can be found in [1] and be downloaded from the URL given below:
Dataset URL: https://drive.grand-challenge.org

Task: To develop an automatic vessel segmentation using neural networks and deep learning which provides high accuracy, robust to varying vessel and background between particular patients, and faster so that it can be deployed in the real world.

Relevant Papers

[1]. Mubbashar, M., Usman, A., & Akram, M. U. (2011, July). Automated system for macula detection in digital retinal images. In 2011 International Conference on Information and Communication Technologies (pp. 1-5). IEEE. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5983555&casa_token=18ETYNnrOBIAAAAA:SluHLvnHoGugqdAysneUUA4Ik6FmR2N0Z3qM9PrWR8Od6ewX30FEL1MHIKCItGUm_NDRs4Y9DA&tag=1

[2]. Gegundez-Arias, M. E., Marin-Santos, D., Perez-Borrero, I., & Vasallo-Vazquez, M. J. (2021). A new deep learning method for blood vessel segmentation in retinal images based on convolutional kernels and modified U-Net model. Computer Methods and Programs in Biomedicine, 205, 106081. https://www.sciencedirect.com/science/article/pii/S0169260721001565?casa_token=hjRsbNxFdggAAAAA:Gvf--glb411tZ40QZwaDmec9SxPL41KxKRuaBFB7sf9MXPXXM9ezB0z2MZPfokZO6txQqbs9Lw

[3]. Chala, M., Nsiri, B., El yousfi Alaoui, M. H., Soulaymani, A., Mokhtari, A., & Benaji, B. (2021). An automatic retinal vessel segmentation approach based on Convolutional Neural Networks. Expert Systems with Applications, 184, 115459. https://www.sciencedirect.com/science/article/pii/S0957417421008721?casa_token=SIfpKyib9i8AAAAA:AUpqSLN_U7GJnGkhqqpRZzUHBb6Fws5cqGFDvsoYnVSod-DD-f5M2w7Y-F4CILfR3fOdyclrkg

[4]. Uysal, E., & Güraksin, G. E. (2021). Computer-aided retinal vessel segmentation in retinal images: convolutional neural networks. Multimedia Tools and Applications, 80, 3505-3528. https://link.springer.com/article/10.1007/s11042-020-09372-w

[5]. Desiani, A., Suprihatin, B., Efriliyanti, F., Arhami, M., & Setyaningsih, E. (2022). VG-DropDNet a robust architecture for blood vessels segmentation on retinal image. IEEE Access, 10, 92067-92083. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9869834

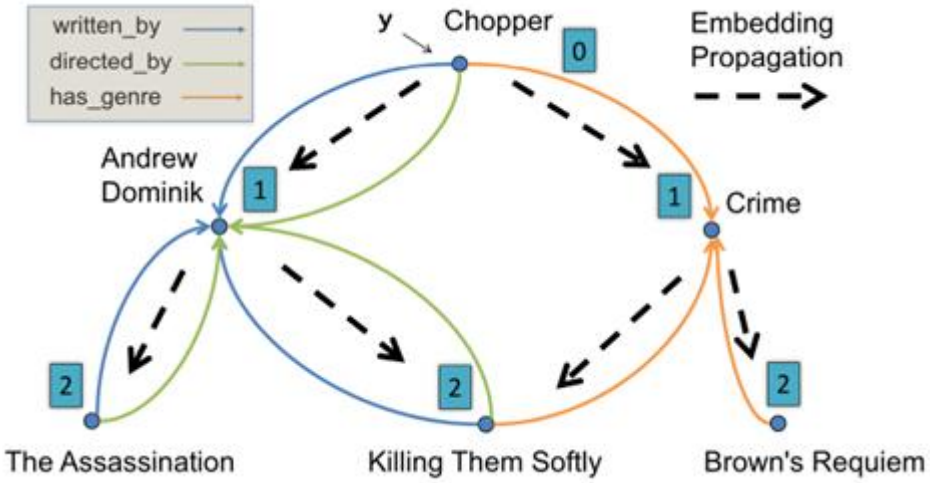
Project ID: 034

Project Title: Question Answering on Movie Knowledge Graphs

Area of Research: Natural Language Processing (NLP).

Problem Statement: Question Answering is the task of automatically generating correct answers to factual questions. The task has numerous practical applications, including the development of Recommender Systems, Digital Assistants and Customer Service systems. This project deals with Question Answering in the Movie Domain: the objective is to train a model on a large Knowledge Graph of movie facts, so that it learns to predict the answer to unseen questions in the test set.

Dataset:



MetaQA (MoviE Text Audio QA) is a dataset of over 400,000 question-answer pairs on the movie domain. Questions in the corpus include identifying the cast of the movie, the director, the genre or the year it was released in.

Dataset URL: https://github.com/yuyuz/MetaQA

Dataset Paper: https://arxiv.org/pdf/1709.04071

Task: Train a model to extract relevant knowledge from graph-structured movie facts, so that it learns to generate the correct response to factual questions about movies.

Relevant Papers

[1].Gupta, Aayushee, et al. "An Inference Approach To Question Answering Over Knowledge Graphs." arXiv preprint arXiv:2112.11070 (2021). https://arxiv.org/pdf/1709.04071

[2]. Saxena, Apoorv, Aditay Tripathi, and Partha Talukdar. "Improving multi-hop question answering over knowledge graphs using knowledge base embeddings." Proceedings of the 58th annual meeting of the association for computational linguistics. 2020. https://aclanthology.org/2020.acl-main.412.pdf

[3]. Sun, Haitian, Tania Bedrax-Weiss, and William Cohen. "PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text." Proceedings of EMNLP-IJCNLP. 2019.. https://arxiv.org/pdf/1904.09537

Project ID: 035

Project title: GAN-Based Semantic Segmentation Enhancement for Complex Scenes in ADE20K

Area of Research: Computer Vision (CV)

Problem Statement:

This project aims to enhance the accuracy and detail of semantic segmentation for complex scenes in the ADE20K dataset using Generative Adversarial Networks (GANs). By leveraging the capabilities of GANs, the project will focus on refining segmentation maps to capture fine-grained details and improve overall segmentation performance. The approach involves training a GAN-based model that learns to generate high-quality segmentation masks from input images, effectively addressing challenges such as occlusions, varying lighting conditions, and diverse scene compositions. The enhanced segmentation maps produced by the GAN will be evaluated against standard benchmarks to demonstrate improvements over traditional segmentation methods.

Dataset:

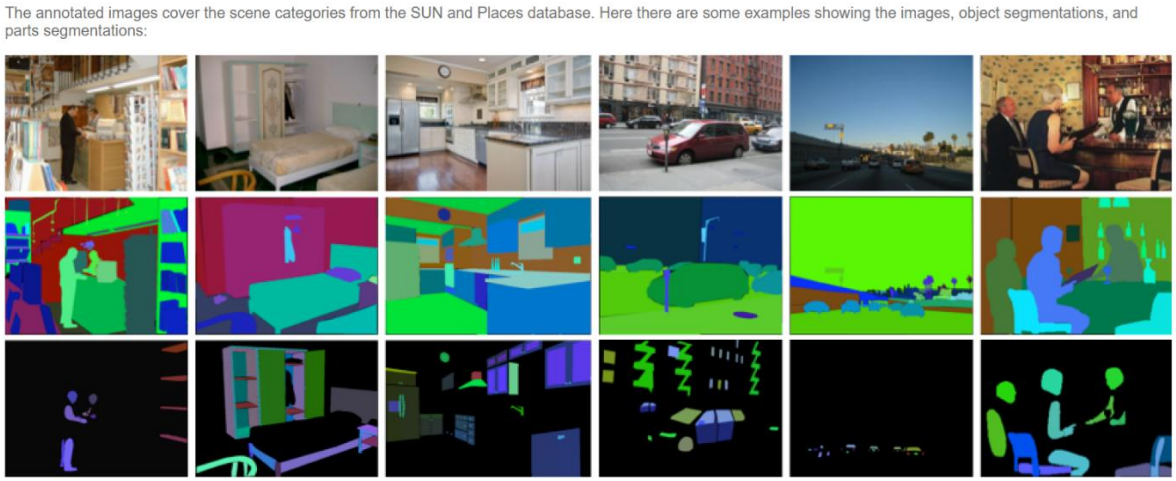


Figure 1: The annotated images cover the scene categories from the SUN and Places database. Here there are some examples showing the images, object segmentations, and parts segmentations.

ADE20K is a dataset for instance segmentation, semantic segmentation, and object detection tasks. It is applicable or relevant across various domains. Also, it is used in the robotics industry.

Dataset URL: https://datasetninja.com/ade20k

https://groups.csail.mit.edu/vision/datasets/ADE20K/

Task: The objective of this project is to develop a GAN-based model to enhance semantic segmentation accuracy for complex scenes in the ADE20K dataset. Traditional segmentation methods often struggle with intricate details and diverse scene elements, leading to suboptimal results. This project will utilize the generative capabilities of GANs to refine and improve segmentation maps, focusing on capturing fine-grained details and handling challenging conditions such as occlusions and varying lighting. By training the GAN on ADE20K, the model will learn to produce high-quality segmentation masks from input images. The effectiveness of the enhanced segmentation will be evaluated through standard metrics, demonstrating its superiority over conventional methods.

Relevant papers:

1. Abdollahi, Abolfazl, et al. "Improving road semantic segmentation using generative adversarial network." IEEE Access 9 (2021): 64381-64392.

2. Zhang, Chaoyi, et al. "MS-GAN: GAN-based semantic segmentation of multiple sclerosis lesions in brain magnetic resonance imaging." 2018 Digital Image Computing: Techniques and Applications (DICTA). IEEE, 2018.

3. Pan, Zhihao, et al. "Automatic pavement crack segmentation using a generative adversarial network (GAN)-based convolutional neural network." Results in Engineering 19 (2023): 101267.

Project ID: 036

Project Title: Dumpsite Monitoring

Area of Research: Computer Vision, Object Detection

Problem Statement: With the advancement of global civilisation, monitoring and managing dumpsites have become essential parts of environmental governance in various countries. Dumpsite locations are difficult to obtain in a timely manner by local government agencies and environmental groups. The World Bank shows that governments need to spend massive labour and economic costs to collect illegal dumpsites to implement management. The primary task of dumpsite monitoring is to regularly confirm their locations, which the environmental department often does at enormous labour cost. The goal of this project is to localize the dumpsite positions from satellite imagery using artificial intelligent algorithms and estimate their categories.

Dataset:

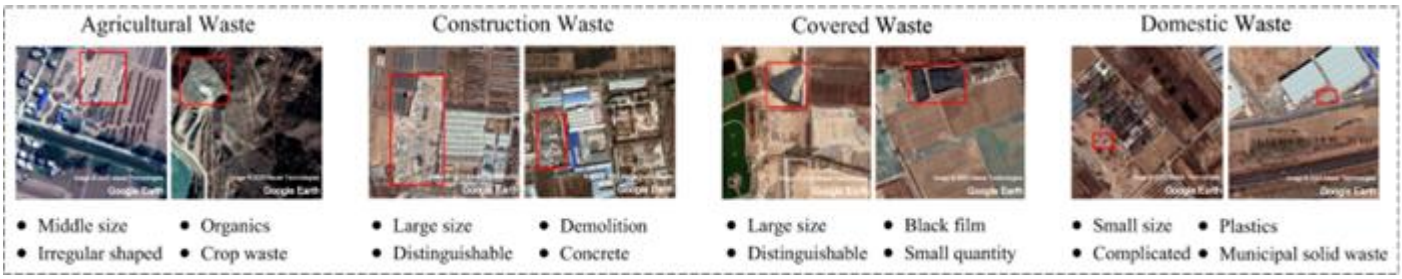


Figure 1 Basic information of dumpsite dataset and typical examples of four categories.

Figure 2 Dataset Snapshot

Dataset URL: https://www.scidb.cn/en/s/6bq2M3

Dataset paper: https://www.nature.com/articles/s41467-023-37136-1

Task: To develop a dumpsite monitoring system using object detection algorithms, including predicting the bounding box of the dumpsite regions and providing their category labels.

Relevant Papers

[1]. Dongshuo Yin. Global Dumpsite Test Data[DS/OL]. V4. Science Data Bank, 2023[2024-05-16]. https://cstr.cn/31253.11.sciencedb.07187. CSTR:31253.11.sciencedb.07187.

[2]. Sun, X., Yin, D., Qin, F. et al. Revealing influencing factors on global waste distribution via deep-learning based dumpsite detection from satellite imagery. Nat Commun 14, 1444 (2023). https://doi.org/10.1038/s41467-023-37136-1

Project ID: 037

Project Title: Pedestrian Trajectory Prediction in Crowds

Area of Research: Trajectory Prediction, Sequence Modelling, Generative Model

Problem Statement: Pedestrian behaviour modelling and analysis is important for crowd scene understanding and has various applications in video surveillance, autonomous vehicles and social robotics. For example, pedestrians follow different trajectories to avoid collisions with neighbours and static obstacles. The goal of this project is to build a pedestrian trajectory prediction system using the GC Dataset with the requirements: **given a 3.2 second of historical trajectory data (2D coordinate sequences) and video frames, the system predicts the future path for the next 4.8 seconds.** To make accurate predictions, there are some advanced directions, including (1) the social interaction modelling among pedestrians, e.g., group detection; (2) the scene interaction modelling between pedestrians and static environment; (3) multi-future trajectory prediction using generative models (e.g., GAN).

Dataset:

Figure 1 Trajectory Prediction (left), Grand Central Station (GC) dataset (mid) and example of predictions (right).

Dataset URL: https://www.dropbox.com/s/7y90xsxq0l0yv8d/cvpr2015_pedestrianWalkingPathDataset.rar?e=1&dl=0
Dataset Paper: https://ieeexplore.ieee.org/document/7298971/

Task: This task is to build a pedestrian trajectory prediction system using the GC Dataset with the basic requirements: **given a 3.2 second of historical trajectory data and video frames, the system makes accurate and socially acceptable predictions for the 4.8 seconds.** Commonly used evaluation metrics are:

- Average Displacement Error: The average L2 distance between the predictions and the ground truth.
- Collision Rate (optional): whether the collisions exist (e.g., distance < 0.2m) among predictions.

Relevant Papers

[1]. B. Zhou, X. Wang and X. Tang. "Understanding Collective Crowd Behaviors: Learning a Mixture Model of Dynamic Pedestrian-Agents." in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[2]. S. Yi, H. Li and X. Wang, "Understanding pedestrian behaviors from stationary crowd groups," 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3488-3496.

[3] Y. Xu, Z. Piao and S. Gao, "Encoding Crowd Interaction with Deep Neural Network for Pedestrian Trajectory Prediction," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5275-5284.

[4] Gupta, Agrim, et al. "Social gan: Socially acceptable trajectories with generative adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Project ID: 038

Project Title: Semantic Segmentation in Unstructured Environments

Area of Research: Computer Vision, Semantic Segmentation, Unstructured Environment

Problem Statement: Semantic scene understanding is an essential technique in autonomous robot navigation in real-world environments. Existing semantic segmentation methods mainly focus mainly on urban cities. However, they mostly ignore elements that may be present in unstructured environments, e.g., different terrains and vegetations in the mountain places. Fortunately, there is an existing dataset named the Robot Unstructured Ground Driving (RUGD) dataset, which is composed of a set of video sequences collected from a small, unmanned ground robot performing an exploration task in a variety of natural, unstructured environments and semi-urban areas. This project aims to explore an accurate, efficient semantic segmentation model that can be used in unstructured environment using RUGD dataset.

Dataset:

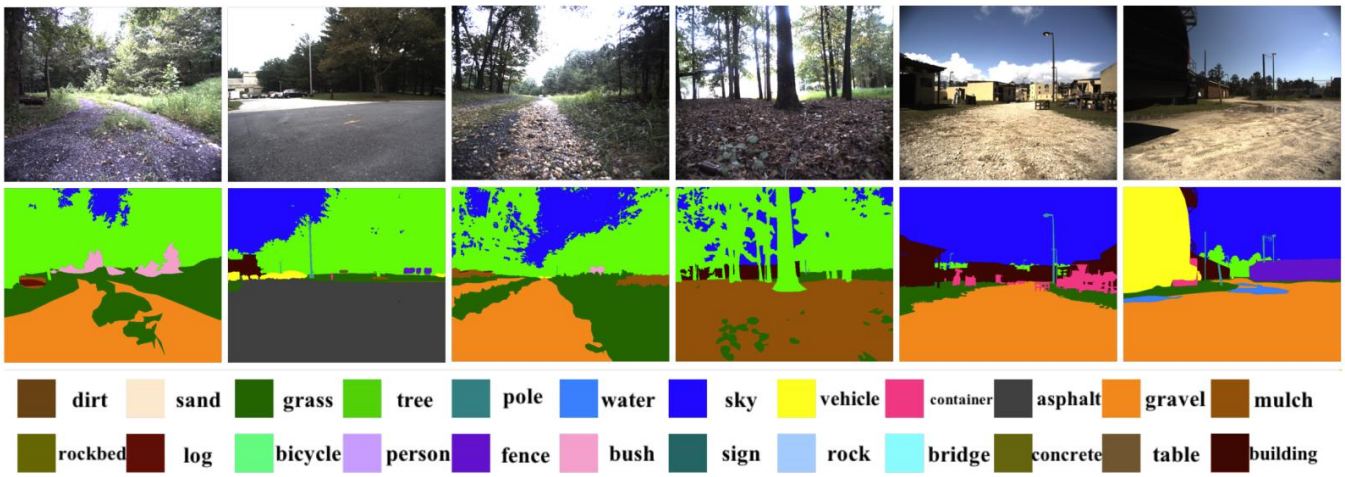


Figure 1 Example ground truth annotations provided in the RUGD dataset. Frames from the video sequences are densely annotated with pixel-wise labels from 24 different visual classes.

Dataset URL: http://rugd.vision/
Dataset paper: https://arxiv.org/html/2404.18750v1

Task: To develop a semantic segmentation model used in unstructured environment using RUGD dataset.

Relevant Papers

[1]. M. Wigness, S. Eum, J. G. Rogers, D. Han and H. Kwon, "A RUGD Dataset for Autonomous Navigation and Visual Perception in Unstructured Outdoor Environments," 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2019, pp. 5000-5007, doi: 10.1109/IROS40897.2019.8968283.

[2]. Guan, T., Kothandaraman, D., Chandra, R., Sathyamoorthy, A.J., Weerakoon, K.M., & Manocha, D. (2021). GA-Nav: Efficient Terrain Segmentation for Robot Navigation in Unstructured Outdoor Environments. IEEE Robotics and Automation Letters, 7, 8138-8145.

[3]. Wijayathunga, L.; Rassau, A.; Chai, D. Challenges and Solutions for Autonomous Ground Robot Scene Understanding and Navigation in Unstructured Outdoor Environments: A Review. Appl. Sci. 2023, 13, 9877. https://doi.org/10.3390/app13179877

Project ID: 039

Project Title: Remote Sensing Image Caption Generation Using the RSICD Dataset

Area of Research: Computer Vision and Natural Language Processing

Problem Statement: Remote sensing images captured by satellites and airplanes contain complex scenes that are difficult to describe with simple labels. Accurate and flexible descriptions of these images are crucial for applications such as environmental monitoring, urban planning, and disaster management. The challenge lies in generating detailed and coherent sentences that capture the semantics of remote sensing images, considering scale, category, and rotation ambiguities. This project aims to develop a deep learning model for generating captions for remote sensing images using the RSICD dataset, advancing the task of remote sensing image captioning.

Dataset:

RSICD (Remote Sensing Image Captioning Dataset) [1]

The RSICD dataset consists of more than 10,000 remote sensing images, each annotated with five descriptive sentences. The images are sourced from platforms like Google Earth, Baidu Map, MapABC, and Tianditu, and are fixed at 224 × 224 pixels. The dataset includes a diverse range of scenes and objects, making it a comprehensive resource for developing and evaluating image captioning models.

1. An old court is surrounded by white houses.
2. A playground is surrounded by many trees and long buildings.
3. A playground with basketball fields next to it is surrounded by many green trees and buildings.
4. Many green trees and several long buildings are around a playground.
5. This narrow, oval football field and closing basketball court, tennis court, parking lot together form this area, with plants wreathing it.

1. Four planes are stopped on the open space between the parking lot.
2. Four white planes are between two white buildings.
3. Some cars and two buildings are near four planes.
4. Four planes are parked next to two buildings on an airport.
5. Four white planes are between two white buildings.

Fig. 1. Example of images and corresponding five sentences each image selected from our data set.

Dataset URL: https://github.com/201528014227051/RSICD_optimal

Tasks: Develop an automated system for generating descriptive captions for remote sensing images using the RSICD dataset. The objectives are:

- To create a model that generates accurate and coherent captions for remote sensing images.
- To address the challenges of scale, category, and rotation ambiguities inherent in remote sensing images.

Relevant Papers

[1]. Lu, X., Wang, B., Zheng, X., & Li, X. (2018). "Exploring Models and Data for Remote Sensing Image Caption Generation." IEEE Transactions on Geoscience and Remote Sensing, 56(4), 2183-2195.

Project ID: 040

Project Title: Real-Time Point Cloud Shape Completion using RL-GAN-Net.

Area of Research: Reinforcement Learning and Generative Adversarial Networks (GANs).

Problem Statement: The problem of point cloud shape completion involves reconstructing missing parts of objects from partial and noisy point cloud data, which is crucial for applications in robotics, augmented reality, and autonomous driving. Traditional methods often struggle with the complexity and variability of shapes. This project aims to develop a real-time shape completion system using (Reinforcement Learning - Generative Adversarial Networks) RL-GAN-Net, a network that leverages reinforcement learning and generative adversarial networks to effectively complete point clouds. The primary objective is to train an agent that can accurately and efficiently reconstruct shapes from incomplete data using synthetic datasets. We recommend you use “latent_3d_points” dataset which a dataset of cloud shape as shown below.

Dataset:

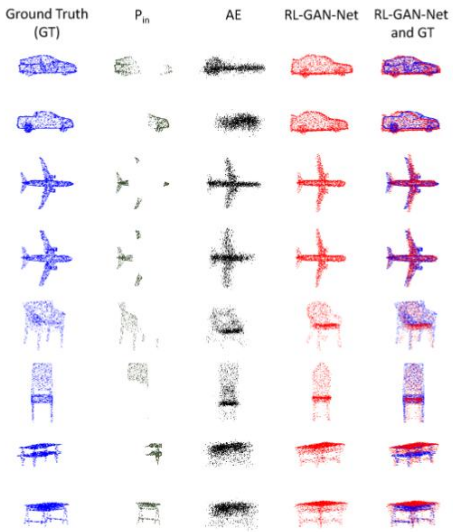


Figure 1. Qualitative results of point cloud shape completion given input data missing 70% of its original points by RL-GAN-Net, which observes a partial input point cloud data (P_{in}) and completes the shape. AE represent autoencoder performance.

Dataset URL: https://github.com/eclipse-sumo/sumo
Dataset paper: Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y. P., Hilbrich, R., ... & Wießner, E. (2018, November). Microscopic traffic simulation using sumo. In 2018 21st international conference on intelligent transportation systems (ITSC) (pp. 2575-2582). IEEE.

Task: Firstly, you should be able to set up the SUMO simulator and configure various traffic scenarios for training the traffic light agent. This includes creating at least a simple traffic pattern with different vehicle densities, arrival rates, and an intersection layout. Then, you need to develop a deep reinforcement learning algorithm to train the traffic light agent. The algorithm should enable the agent to learn optimal traffic signal policies from its interactions with the simulated environment. The trained agent must manage traffic lights efficiently in various test scenarios, minimizing average waiting times, congestion, and improving overall traffic flow.

Relevant Papers:

[1]. Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y. P., Hilbrich, R., ... & Wießner, E. (2018, November). Microscopic traffic simulation using sumo. In 2018 21st international conference on intelligent transportation systems (ITSC) (pp. 2575-2582). IEEE.
[2]. Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529-533.
[3]. Van der Pol, Elise, and Frans A. Oliehoek. "Coordinated deep reinforcement learning for traffic light control." NeurIPS Workshop on Modeling and Decision-making in the Spatiotemporal Domain. *2016.*
[4]. Guo, M., Wang, P., Chan, C. Y., & Askary, S. (2019, October). A reinforcement learning approach for intelligent traffic signal control at urban intersections. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 4242-4247). IEEE.

Project ID: 041

Project Title: MS Lesion Segmentation in Brain MRI

Area of Research: Computer Vision (Segmentation and localization).

Problem Statement: Accurate segmentation of Multiple Sclerosis (MS) lesions in brain MRI images is critical for diagnosis, treatment planning, and monitoring disease progression. Manual delineation of lesions is time-consuming and prone to variability among radiologists. This project aims to develop a deep learning-based system for automated MS lesion segmentation. By leveraging advanced neural network architectures, the system will accurately identify and delineate MS lesions from brain MRI scans, providing clinicians with precise and consistent lesion maps. The proposed solution seeks to improve diagnostic efficiency, treatment decision-making, and patient care in MS management. To do this project, we recommend to you the ISBI 2015 Longitudinal Multiple Sclerosis Lesion Segmentation Challenge dataset comprises longitudinal MRI scans from multiple sclerosis patients, providing a valuable resource for developing algorithms that can accurately segment and track lesions over time.

Dataset:

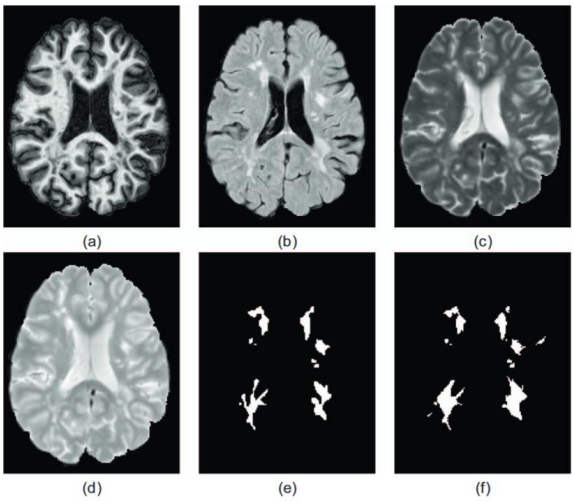


Figure 1. Shown are the preprocessed images from the dataset (a) MPRAGE, (b) FLAIR, (c) T2 -w, and (d) PD-w images for a single time-point from one of the provided Training Set subjects. The corresponding lesion areas by our two raters (doctors) are shown in (e) for Rater #1 and (f) for Rater #2.

Dataset URL: https://smart-stats-tools.org/lesion-challenge
Dataset paper: Carass, Aaron, Snehashis Roy, Amod Jog, Jennifer L. Cuzzocreo, Elizabeth Magrath, Adrian Gherman, Julia Button et al. "Longitudinal multiple sclerosis lesion segmentation: resource and challenge." NeuroImage 148 (2017): 77-102.

Task: To develop an automatic MS lesion segmentation system using neural networks and deep learning which provides high accuracy, Dice Similarity Coefficient (DSC), and low error rate so that it can be deployed in the real world.

Relevant Papers

[1]. Carass, Aaron, Snehashis Roy, Amod Jog, Jennifer L. Cuzzocreo, Elizabeth Magrath, Adrian Gherman, Julia Button et al. "Longitudinal multiple sclerosis lesion segmentation: resource and challenge." NeuroImage 148 (2017): 77-102.
[2]. Noori, Mehrdad, Ali Bahri, and Karim Mohammadi. "Attention-guided version of 2D UNet for automatic brain tumor segmentation." In 2019 9th international conference on computer and knowledge engineering (ICCKE), pp. 269-275. IEEE, 2019.
[3]. Hashemi, Maryam, Mahsa Akhbari, and Christian Jutten. "Delve into multiple sclerosis (MS) lesion exploration: a modified attention U-net for MS lesion segmentation in brain MRI." Computers in Biology and Medicine 145 (2022): 105402.
[4]. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pp. 234-241. Springer International Publishing, 2015.
[5]. Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori et al. "Attention u-net: Learning where to look for the pancreas." arXiv preprint arXiv:1804.03999 (2018).
Project ID: 0041

Project ID: 042

Project Title: Object Detection for Autonomous Drones.

Area of Research: Computer Vision (Detection and recognition).

Problem Statement: Object detection is a hot topic with various applications in computer vision, such as image understanding, autonomous driving, and video surveillance. Object detection on the drone platform is a challenging task due to various factors such as viewpoint changes, occlusion, and scale variations. Traditional methods of object detection struggle with the challenges posed by varying altitudes, speeds, and environmental conditions. This project aims to develop a deep learning-based system for object detection for autonomous drones. By leveraging advanced neural network architectures, the system will accurately identify and classify various objects within images provided by drones. For this project, we recommend using the VisDrone dataset, which provides a large collection of images captured by drones in various scenarios, including urban and rural areas. This dataset includes annotations for multiple object categories, such as cars, pedestrians, bicycles, animals, and trees.

Dataset:

Figure 1. Some annotated example images of the object detection in images task. The dashed bounding box indicates the object is occluded. Different bounding box colors indicate different classes of objects. For better visualization, only some attributes have been displayed.

Dataset URL: https://github.com/VisDrone/VisDrone-Dataset?tab=readme-ov-file

Dataset paper: Zhu, Pengfei, Longyin Wen, Dawei Du, Xiao Bian, Haibin Ling, Qinghua Hu, Qinqin Nie et al. "Visdrone-det2018: The vision meets drone object detection in image challenge results." In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp. 0-0. 2018.

Task: To develop an automatic object detection system using neural networks and deep learning that provides high accuracy, precision, recall, and low error rates, ensuring its suitability for deployment in autonomous drones.

Relevant Papers:

[1]. Zhu, Pengfei, Longyin Wen, Dawei Du, Xiao Bian, Haibin Ling, Qinghua Hu, Qinqin Nie et al. "Visdrone-det2018: The vision meets drone object detection in image challenge results." In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pp. 0-0. 2018.

[2]. Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[3]. Sun, Wei, Liang Dai, Xiaorui Zhang, Pengshuai Chang, and Xiaozheng He. "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring." Applied Intelligence (2022): 1-16.

[4]. Du, Dawei, Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Lin, Qinghua Hu, Tao Peng et al. "VisDrone-DET2019: The vision meets drone object detection in image challenge results." In Proceedings of the IEEE/CVF international conference on computer vision workshops, pp. 0-0. 2019.

Project ID: 043

Project Title: Images Style Transfer

Area of Research: Computer Vision (CV)

Problem Statement:
The objective of image style transfer is to apply artistic features from a style reference image to a target image while preserving the original content. Traditional neural style transfer methods, which rely on convolutional neural networks (CNNs), often struggle with capturing and maintaining the global information of input images due to the inherent locality of CNNs. This results in biased content representation, limiting the effectiveness of these methods.

Nowadays, Generative AI (GenAI) and Generative Adversarial Networks (GANs) offer numerous solution methods to tackle the challenges of image style transfer. One promising direction is the use of advanced models that can better capture global dependencies within images compared to traditional CNNs.

To address these challenges, it is crucial to explore innovative approaches that integrate both local and global image information effectively. Furthermore, existing positional encoding methods often fall short in image style transfer tasks, highlighting the need for more suitable and scale-invariant positional encoding techniques.

By leveraging the potential of GenAI and GANs and addressing the limitations of traditional methods, it is possible to develop more robust solutions for image style transfer. These advancements can significantly improve the accuracy of content preservation and the quality of style application, setting new benchmarks in the field. Qualitative and quantitative experiments are essential to validate the effectiveness of these innovative approaches compared to state-of-the-art methods.

Dataset:

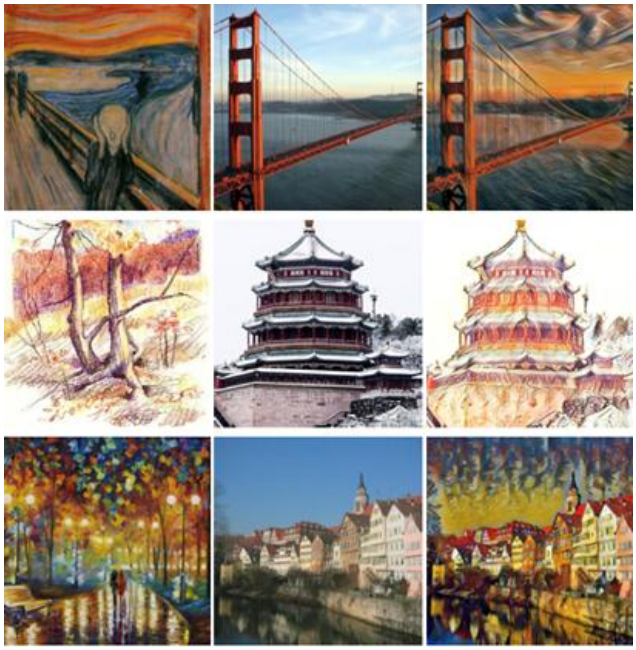


Figure 1. Example images of the COCO dataset are presented. The first column displays the style images, the second column shows the content images, and the results after the style transfer are exhibited in the last column.

Dataset URL: The MS COCO (Microsoft Common Objects in Context) dataset [1] is a large-scale dataset used for object detection, segmentation, key-point detection, and captioning tasks. It comprises 164,000 images and is widely used in computer vision research. More information about the dataset can be found in [1], and it can be downloaded from the following URL:

Style Images URL: https://www.wikiart.org
Content Dataset URL: https://www.kaggle.com/datasets/jeffaudi/coco-2014-dataset-for-yolov3

Task: To develop an automatic image style transfer method using Generative AI (GenAI) or Generative Adversarial Networks (GANs) that provides high accuracy and robustness in transferring artistic styles to images while preserving the original content. This advancement aims to enhance the reliability and applicability of image processing in various real-world applications, including art, design, and media production.

Relevant Papers

[1]. Deng, Y., Tang, F., Dong, W., Ma, C., Pan, X., Wang, L., & Xu, C. (2022). Stytr2: Image style transfer with transformers. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11326-11336).
https://openaccess.thecvf.com/content/CVPR2022/papers/Deng_StyTr2_Image_Style_Transfer_With_Transformers_CVPR_2022_paper.pdf

[2]. Xu, W., Long, C., Wang, R., & Wang, G. (2021). Drb-gan: A dynamic resblock generative adversarial network for artistic style transfer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6383-6392).
https://openaccess.thecvf.com/content/ICCV2021/papers/Xu_DRB-GAN_A_Dynamic_ResBlock_Generative_Adversarial_Network_for_Artistic_Style_ICCV_2021_paper.pdf

[3]. Chen, X., Xu, C., Yang, X., Song, L., & Tao, D. (2018). Gated-gan: Adversarial gated networks for multi-collection style transfer. IEEE Transactions on Image Processing, 28(2), 546-560.
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8463508&casa_token=V0dFwlFXi08AAAAA:35OYqS6AMf9bGBXZJn3tL9RcJSWuGUMC5Bs33FIuxm8MKJAcQ2Nw7WU7h9Q97TGFHifcVKVnJg&tag=1

[4]. Liu, H., Michelini, P. N., & Zhu, D. (2018, August). Artsy-GAN: A style transfer system with improved quality, diversity and performance. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 79-84). IEEE.
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8546172&casa_token=sD9AT9FfY8AAAAAA:x_2LC0tK73HdjihYYkyx7oNxMweu34lrD0goxGgISPDuRJzpNZtZACXBUYsXpBOI5QHOukovog

[5]. Xu, W., Long, C., & Nie, Y. (2023). Learning dynamic style kernels for artistic style transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10083-10092).
https://openaccess.thecvf.com/content/CVPR2023/papers/Xu_Learning_Dynamic_Style_Kernels_for_Artistic_Style_Transfer_CVPR_2023_paper.pdf

Project ID: 044

Project Title: Traffic Light Control Using a Deep Learning Agent.

Area of Research: Reinforcement Learning and Deep Learning.

Problem Statement: The problem of traffic light control involves optimizing the traffic flow at intersections to reduce congestion, waiting times, and emissions. Traditional methods rely on pre-set timers or simple reactive policies, which often fail to adapt to dynamic traffic conditions. This project aims to develop a deep reinforcement learning-based system for intelligent traffic light control, focusing on a simplified urban scenario. The primary objective is to train an agent to manage traffic signals efficiently using real-time traffic data. The project will utilize the SUMO (Simulation of Urban MObility) simulator, providing a realistic and configurable environment for training and testing.

Dataset:

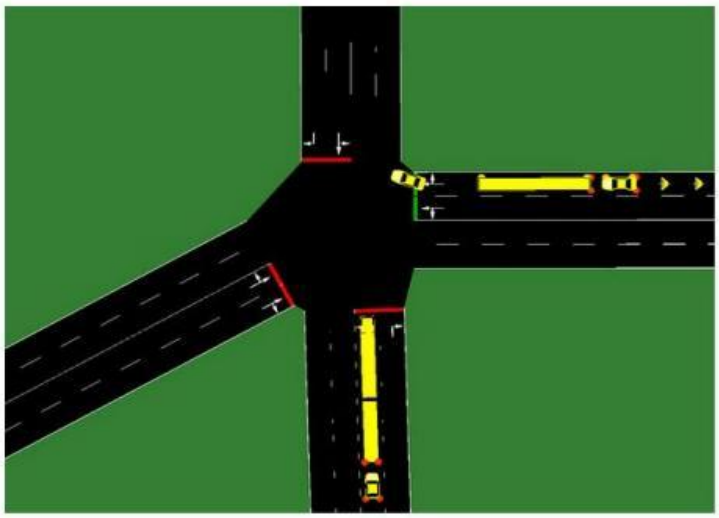


Figure 1. An example of SUMO traffic simulator for a simple environment.

Dataset URL: https://github.com/eclipse-sumo/sumo

Dataset paper: Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y. P., Hilbrich, R., ... & Wießner, E. (2018, November). Microscopic traffic simulation using sumo. In 2018 21st international conference on intelligent transportation systems (ITSC) (pp. 2575-2582). IEEE.

Task: Firstly, you should be able to set up the SUMO simulator and configure various traffic scenarios for training the traffic light agent. This includes creating at least a simple traffic pattern with different vehicle densities, arrival rates, and an intersection layout. Then, you need to develop a deep reinforcement learning algorithm to train the traffic light agent. The algorithm should enable the agent to learn optimal traffic signal policies from its interactions with the simulated environment. The trained agent must manage traffic lights efficiently in various test scenarios, minimizing average waiting times, congestion, and improving overall traffic flow.

Relevant Papers:

[1]. Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y. P., Hilbrich, R., ... & Wießner, E. (2018, November). Microscopic traffic simulation using sumo. In 2018 21st international conference on intelligent transportation systems (ITSC) (pp. 2575-2582). IEEE.

[2]. Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529-533.

[3]. Van der Pol, Elise, and Frans A. Oliehoek. "Coordinated deep reinforcement learning for traffic light control." NeurIPS Workshop on Modeling and Decision-making in the Spatiotemporal Domain. 2016.

[4]. Guo, M., Wang, P., Chan, C. Y., & Askary, S. (2019, October). A reinforcement learning approach for intelligent traffic signal control at urban intersections. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 4242-4247). IEEE.

Project ID: 045

Project Title: Environmental Microorganism Image Analysis Using Deep Learning

Area of Research: Environmental Microbiology and Computer Vision

Problem Statement: Environmental microorganisms (EMs) are ubiquitous and significantly impact human survival and development. The detection and identification of these microorganisms are crucial for various applications, including environmental monitoring, disease prevention, and industrial processes. Traditional methods of identifying EMs involve manual microscopic analysis, which is labor-intensive, time-consuming, and prone to human error. This project aims to develop an automated system for analyzing EM images using deep learning techniques. The goal is to enhance the accuracy and efficiency of tasks such as image denoising, segmentation, feature extraction, classification, and object detection.

Dataset: Environmental Microorganism Image Dataset Sixth Version (EMDS-6) [1]

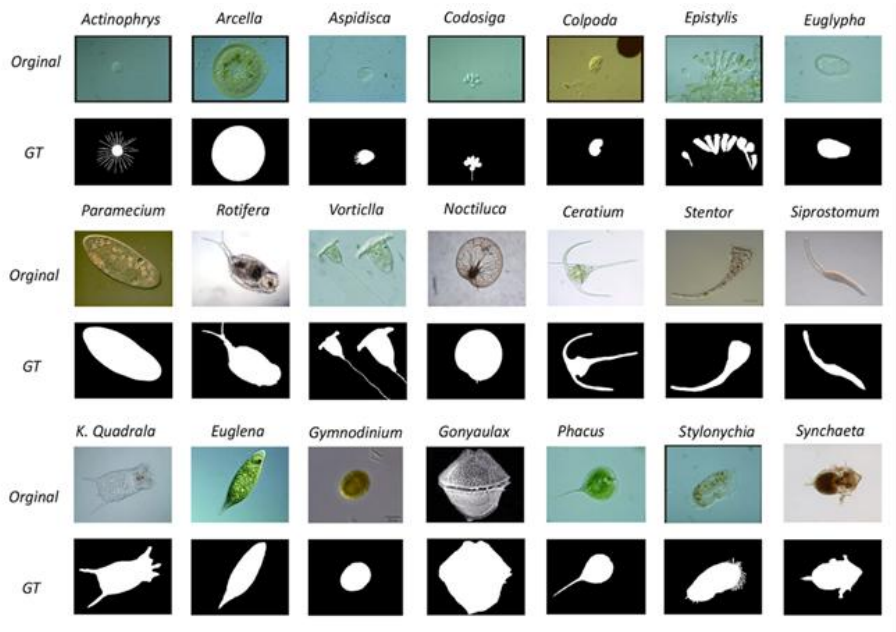


FIGURE 1 | An example of EMDS-6, including original images and GT images

The EMDS-6 dataset contains 21 types of EMs, with each type having 40 original images and 40 corresponding ground truth (GT) images, totaling 1680 images. The dataset is designed to facilitate various image processing and analysis tasks.

Dataset URL: https://figshare.com/articles/dataset/EMDS6/17125025/1

Tasks: Develop an automated system for analyzing environmental microorganism images using deep learning methods. The objectives include:

- Image denoising to reduce noise and enhance image quality.
- Image segmentation to accurately delineate the regions of interest.
- Feature extraction to identify relevant characteristics for classification.
- Image classification to categorize images into different types of EMs.

Relevant Paper:

[1]. Zhao, P., Li, C., Rahaman, M. M., et al. (2022). "EMDS-6: Environmental Microorganism Image Dataset Sixth Version for Image Denoising, Segmentation, Feature Extraction, Classification, and Detection Method Evaluation." Frontiers in Microbiology, 13, 829027.

Project ID: 046

Project Title: Colorectal Histopathological Image Classification Using Deep Learning

Area of Research: Computer Vision

Problem Statement: Colorectal cancer is the third most common cancer worldwide, accounting for approximately 10% of cancer patients. Early detection of the disease is crucial for effective treatment and improved patient outcomes. Histopathological examination is the gold standard for diagnosing colorectal cancer. However, the lack of large, publicly available histopathological image datasets, particularly those derived from enteroscope biopsies, poses a significant challenge for the development and evaluation of automated image classification systems. This project aims to develop an automated system for classifying colorectal histopathological images into various stages of tumor differentiation using advanced deep learning algorithms. The goal is to achieve high accuracy and robustness, making the system suitable for clinical application.

Dataset:: Enteroscope Biopsy Histopathological H&E Image Dataset (EBHI)

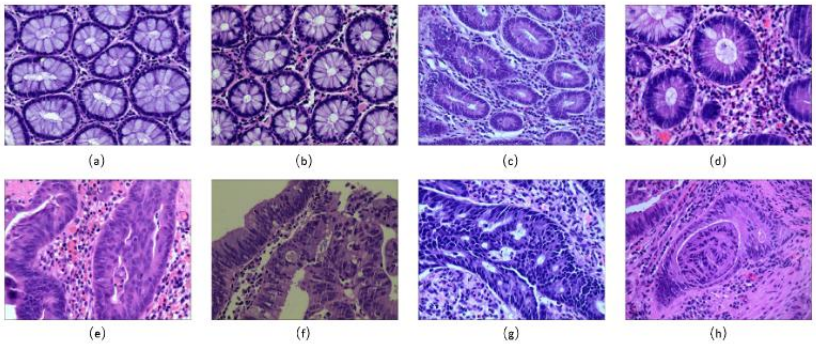


Fig. 1. An example of histopathological images EBHI: (a) Normal, (b) Polyp, (c) (d) Low-grade Intraepithelial Neoplasia, (e) (f) High-grade Intraepithelial Neoplasia, (g) (h) Adenocarcinoma.

The EBHI dataset is a newly published dataset specifically designed for colorectal histopathological image classification. It contains 5532 images with four magnifications (40×, 100×, 200×, and 400×) and five categories of tumor differentiation stages: Normal, Polyp, Low-grade Intraepithelial Neoplasia (Low-grade IN), High-grade Intraepithelial Neoplasia (High-grade IN), and Adenocarcinoma.

Dataset URL: https://figshare.com/articles/dataset/EBH-HE-IDS/16999363/1

Tasks:
Develop an automated colorectal histopathological image classification system using deep learning techniques. The system should classify images into five categories based on tumor differentiation stages. The objectives are:

- To achieve high classification accuracy.
- To ensure robustness to variations in image quality and magnification.
- To evaluate the performance of various deep learning models and compare them.

Relevant Papers

[1]. Hu, W., Li, C., Rahaman, M.M., et al. "EBHI: A new Enteroscope Biopsy Histopathological H&E Image Dataset for image classification evaluation," Physica Medica, 2023.

Project ID: 047

Project Title: Automated Question Answering on News Stories

Area of Research: Natural Language Processing (NLP).

Problem Statement: Question Answering is the task of automatically generating correct answers to factual questions. The task has numerous practical applications, including the development of Recommender Systems, Digital Assistants and Customer Service systems. This project deals with answering questions on news stories from the CNN.

Dataset:

NewsQA is a crowd-sourced dataset consisting of 120,000 Question-Answer pairs on CNN news articles

Dataset URL: https://www.microsoft.com/en-us/research/project/newsqa-dataset/download/

Dataset Paper: https://www.microsoft.com/en-us/research/publication/newsqa-machine-comprehension-dataset/

Task: Given a News Story and a Question, identify a span of text in the news story that correctly answers the question.

Relevant Papers

[1].Tay, Yi, et al. "Densely connected attention propagation for reading comprehension." Advances in neural information processing systems 31 (2018). https://proceedings.neurips.cc/paper_files/paper/2018/file/7b66b4fd401a271a1c7224027ce111bc-Paper.pdf

[2]. Yasunaga, Michihiro, Jure Leskovec, and Percy Liang. "LinkBERT: Pretraining Language Models with Document Links." ICML 2022 2nd AI for Science Workshop. 2022. https://arxiv.org/pdf/2203.15827

[3]. Joshi, Mandar, et al. "Spanbert: Improving pre-training by representing and predicting spans." Transactions of the association for computational linguistics 8 (2020): 64-77. https://arxiv.org/pdf/1907.10529v3

Project ID: 048

Project Title: Cattle Detection and Counting in UAV Images Dataset

Area of Research: Computer Vision (CV).

Problem Statement: Cattle detection and counting are essential tasks in modern agriculture to ensure effective livestock management and optimize farm productivity. Manual counting and monitoring are not only labor-intensive and time-consuming but also highly prone to inaccuracies, especially in large-scale farming operations. This research aims to develop an automated system for detecting and counting cattle using UAV (Unmanned Aerial Vehicle) images. The system should employ advanced artificial intelligence techniques to provide accurate, real-time monitoring and counting of cattle, thereby enabling efficient livestock management.

Dataset:

Figure 1: Example of cattle images form Unmanned Aerial Vehicle (UAV) data samples

Dataset URL: https://datasetninja.com/cattle-detection-and-counting-in-uav-images

Dataset paper:

Shao, Wen, et al. "Cattle detection and counting in UAV images based on convolutional neural networks." *International Journal of Remote Sensing* 41.1 (2020): 31-52.

Task: The primary task is to develop an automated cattle detection and counting system using deep learning techniques for modern farming. This system should handle variations in appearance due to cattle orientations, and other environmental factors. It involves collecting of UAV images, applying preprocessing and data augmentation techniques (if needed), and developing deep learning models, including leveraging pre-trained models and designing custom model. The task also includes integrating object detection with counting mechanisms, training and validating the models using cross-validation, and evaluating performance with relevant metrics. Additionally, the model should be optimized for speed and efficiency to enable real-time deployment with a user-friendly interface for real-time monitoring and field testing to ensure robustness and usability in modern farming environments.

Relevant Papers

[1]. YuanQiang, Cai, et al. "Guided attention network for object detection and counting on drones." *Proceedings of the 28th ACM international conference on multimedia*. 2020.
[2]. Sarwar, Farah, et al. "Detecting and counting sheep with a convolutional neural network." *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2018.
[3]. Xu, Beibei, et al. "Livestock classification and counting in quadcopter aerial images using Mask R-CNN." *International Journal of Remote Sensing* 41.21 (2020): 8121-8142.
[4]. Kellenberger, Benjamin, Michele Volpi, and Devis Tuia. "Fast animal detection in UAV images using convolutional neural networks." *2017 IEEE international geoscience and remote sensing symposium (IGARSS)*. IEEE, 2017.
[5]. Xu, Beibei, et al. "Automated cattle counting using Mask R-CNN in quadcopter vision system." *Computers and Electronics in Agriculture* 171 (2020): 105300.

Project ID: 049

Project Title: Leaf Disease Classification

Area of Research: Computer Vision (CV).

Problem Statement: Leaf disease classification is an important task in agriculture because it helps with early detection and control of illnesses that can have a significant influence on crop output and quality. Manual leaf disease inspection is labor-intensive, time-consuming, and prone to inaccuracy, particularly in large-scale farming operations. The purpose of this research is to create an automated method for classifying leaf diseases that employs advanced artificial intelligence techniques. This system should be very accurate, resilient to environmental fluctuations, and capable of real-time monitoring and detection to allow for rapid intervention and control actions.

Dataset:

Figure 1: Example of different leaf disease image data samples

Dataset URL: https://data.mendeley.com/datasets/tywbtsjrjv/1

Dataset paper: Geetharamani, G., and Arun Pandian. "Identification of plant leaf diseases using a nine-layer deep convolutional neural network." *Computers & Electrical Engineering* 76 (2019): 323-338.

Task: The primary task is to develop a neural network-based leaf disease classification system that can accurately identify and classify different types of leaf diseases. The system should be able to handle variations in appearance due to different lighting conditions, leaf orientations, and other environmental factors. Additionally, the model should be optimized for speed to enable real-time deployment in agricultural settings.

Relevant Papers

[1]. Barbedo, Jayme Garcia Arnal. "Plant disease identification from individual lesions and spots using deep learning." *Biosystems engineering* 180 (2019): 96-107.
[2]. Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection." *Frontiers in plant science* 7 (2016): 215232.
[3]. Sun, Guiling, Xinglong Jia, and Tianyu Geng. "Plant diseases recognition based on image processing technology." *Journal of Electrical and Computer Engineering* 2018 (2018).
[4]. Ferentinos, Konstantinos P. "Deep learning models for plant disease detection and diagnosis." *Computers and electronics in agriculture* 145 (2018): 311-318.
[5]. Sladojevic, Srdjan, et al. "Deep neural networks based recognition of plant diseases by leaf image classification." *Computational intelligence and neuroscience* 2016 (2016).

Project ID: 050

Project Title: Abnormal Tooth Detection with Dental Enumeration and Diagnosis Using Deep Learning on Panoramic X-rays

Area of Research: Computer Vision, Medical Image Analysis, Dental Imaging

Problem Statement: For efficient treatment planning and to reduce procedural errors in dental practice, it is essential to accurately detect and diagnose aberrant teeth in dental X-rays. On the other hand, manual panoramic X-ray analysis can be laborious and prone to human error. Using the Fédération Dentaire Internationale (FDI) system, the project aims to create a deep learning-based system that can automatically identify anomalous teeth, count them, and provide related diagnoses including caries, deep caries, periapical lesions, and impacted teeth. This research seeks to increase the effectiveness and precision of dental anomaly detection, which will ultimately enhance patient care and outcomes in dental practice. It does this by utilizing sophisticated deep learning algorithms and the hierarchically annotated DENTEX dataset.

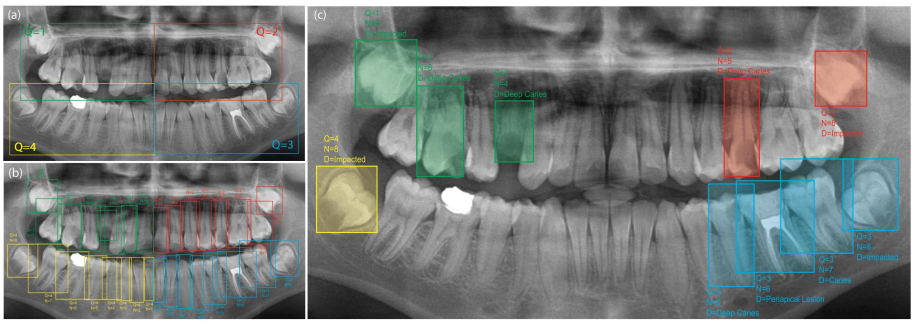


Figure 1. Example panoramic dental X-ray from the DENTEX dataset with annotated abnormal teeth (c).

For this project you need to use only label (c). DO NOT USE label (a) and label (b) and unlabeled data

Dataset URL: https://huggingface.co/datasets/ibrahimhamamci/DENTEX

Task: Develop a deep learning-based model for automatic detection of abnormal teeth in panoramic dental X-rays using the DENTEX dataset. Utilize the hierarchically annotated data to train models for abnormal tooth detection with associated diagnoses. Explore state-of-the-art object detection and multi-label classification architectures, such as Faster R-CNN, YOLO (v7 or v8), or custom-designed models. Employ data augmentation techniques to handle variations in image quality and tooth appearance. Evaluate the model's performance using metrics such as average precision (AP), average recall (AR), and F1 score for each hierarchical level. Compare the model's performance with existing dental abnormality detection methods and assess its potential for integration into clinical workflows.

Relevant Papers:

[1]. Hamamci, I. E., Er, S., Simsar, E., Yuksel, A. E., Gultekin, S., Ozdemir, S. D., ... & others. (2023). DENTEX: An Abnormal Tooth Detection with Dental Enumeration and Diagnosis Benchmark for Panoramic X-rays. arXiv preprint arXiv:2305.19112. https://arxiv.org/abs/2305.19112

[2]. Hamamci, I. E., Er, S., Simsar, E., Sekuboyina, A., Gundogar, M., Stadlinger, B., ... & Menze, B. (2023). Diffusion-based hierarchical multi-label object detection to analyze panoramic dental x-rays. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 389-399). Springer, Cham. https://arxiv.org/abs/2303.06500

Project ID: 051

Project Title: Breast Cancer Classification and Segmentation Using Deep Learning on Ultrasound Images

Area of Research: Computer Vision, Medical Image Analysis, Oncology

Problem Statement: Breast cancer is one of the leading causes of death among women worldwide. Early detection plays a crucial role in reducing mortality rates and improving patient outcomes. Ultrasound imaging is a safe and widely used technique for breast cancer examination and early detection. However, the interpretation of ultrasound images can be challenging, even for experienced radiologists, due to the complexity and variability of breast lesions. The goal of this project is to develop a deep learning-based system for accurate classification and segmentation of breast masses in ultrasound images. By leveraging state-of-the-art deep learning techniques and a comprehensive dataset of breast ultrasound images, this project aims to assist radiologists in the early detection and diagnosis of breast cancer, potentially improving the efficiency and accuracy of the screening process.

Dataset:

Figure 1. Samples of ultrasound breast images from the dataset, categorized into normal, benign, and malignant cases.

The Breast Ultrasound Dataset consists of 780 ultrasound images of breast masses, categorized into three classes: normal, benign, and malignant. The dataset was collected from 600 female patients aged between 25 and 75 years old at Baheya Hospital for Early Detection & Treatment of Women's Cancer in Cairo, Egypt. The images have an average size of 500 × 500 pixels and are in PNG format. In addition to the original ultrasound images, the dataset also includes corresponding ground truth mask images for each lesion, enabling the development and evaluation of segmentation models.

Dataset URL: https://huggingface.co/datasets/gymprathap/Breast-Cancer-Ultrasound-Images-Dataset

Task: Develop a deep learning-based pipeline for breast cancer classification and segmentation using the Breast Ultrasound Dataset. Preprocess the ultrasound images, apply data augmentation techniques, and split the dataset into training, validation, and testing subsets. For the classification task, explore state-of-the-art deep learning architectures such as ResNet, DenseNet, or EfficientNet, and train a model to classify the images into normal, benign, and malignant categories. For the segmentation task, utilize semantic segmentation architectures like U-Net, DeepLab, or FCN, and train a model to accurately delineate the boundaries of breast masses in the ultrasound images.

Relevant Papers:

[1]. Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. Data in Brief, 28, 104863. https://www.sciencedirect.com/science/article/pii/S2352340919312181

[2]. Yap, M. H., et al. (2018). Automated breast ultrasound lesions detection using convolutional neural networks. IEEE Journal of Biomedical and Health Informatics, 22(4), 1218-1226. https://ieeexplore.ieee.org/document/8003418

[3]. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241). Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28

Project ID: 052

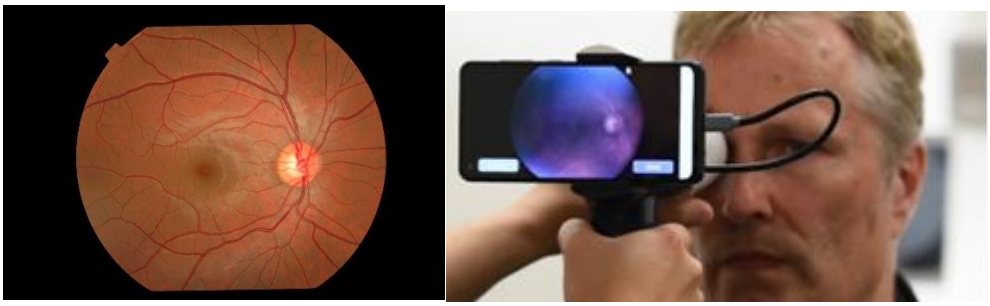
Project Title: Development of a retinal image-based biometric identification system using deep learning

Area of Research: Computer Vision

Problem Statement: Biometric identification and authentication have become crucial in today's digitalised world. Although the conventional biometrics systems such as fingerprints and facial detection are widely being used, they have failed to provide reliability due to several pitfalls. Given the unique nature of human eye, iris-based identification is widely being recognized for high security applications. But it suffers from several limitations, such as, varied pupil size, requirement of proper alignment and positioning and specialised devices. However, recent studies have found that retinal imaging shows promise and reliability in human identification with better accuracy. Due to the development of mobile camera-based retinal imaging/ mobile fundoscopy, retinal imaging based biometric identification can open a new window to the biometric security—being not only a cheap method but also adding another layer to the retinal image-based security system.

This study will aim at developing deep learning methods for identifying individual based on retinal images. Multiple datasets can be utilized, including Retina Identification Database (RIDB) is comprised of Retinal Fundus Images captured using the TOPCON-TRC camera. The dataset contains 100 images with 1504 x 1000 resolution compressed in JPEG format, captured from 20 individuals (five samples per person) with no retinal disease. In this study, RIDB dataset will be used for training and testing of a retinal recognition system. The outcome of this project will help to develop a robust system for biometric identification and authentication for high security applications.

Details of the dataset is provided in Dataset URL.
Dataset:



(a) (b)

Figure 1. Examples of (a) Fundus retinal image (b) human identification/ image capture

Dataset URL: Retina Identification Database (RIDB) - Mendeley Data

Task: Develop a retinal image-based biometric identification system using deep learning

Relevant Papers:

[1]. Akram, M. U., Salam, A. A., Khawaja, S. G., Naqvi, S. G. H., & Khan, S. A. (2020). RIDB: a dataset of fundus images for retina based person identification. *Data in Brief*, *33*, 106433.

https://doi.org/10.1016/j.dib.2020.106433

[2]. Waheed, Z., Waheed, A., & Akram, M. U. (2016, January). A robust non-vascular retina recognition system using structural features of retinal image. In *2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)* (pp. 101-105). IEEE.
https://doi.org/10.1016/j.dib.2020.106433

Project ID: 053

Project Title: Retinal Image analysis using deep learning for multi-disease detection system

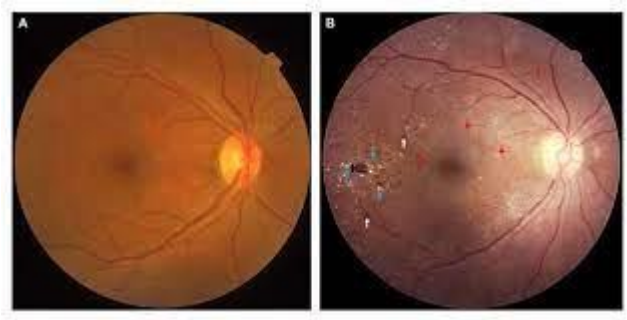
Area of Research: Computer Vision

Problem Statement: According to the World Health Organization's world report on vision 2019, there are 2.2 billion visually impaired persons worldwide, with at least 1 billion having a vision impairment that could have been avoided or that has yet to be addressed. In terms of eye care, the world faces significant challenges, including disparities in coverage and quality of prevention, treatment, and rehabilitation services. Visual impairment could be avoided if ocular disorders were detected and diagnosed early. Most of the studies using colour fundus images have a primary focus on diabetic retinopathy, glaucoma, and age-related macular degeneration, as well as a few other common disorders. However, one barrier to ophthalmologists using a computer-aided diagnosis tool is that sight-threatening uncommon illnesses, such as, central retinal artery blockage or anterior ischemic optic neuropathy are frequently overlooked. As such, an automated computer aided tool is crucial for the diagnosis of common as well as the rear diseases.

The aim of this project is to to develop methods for automatic ocular disease classification of frequent diseases along with the rare pathologies. For this purpose, Retinal Fundus Multi-disease Image Dataset (RFMiD) will be utilised which is the only publicly available dataset that constitutes such a wide variety of diseases that appear in routine clinical settings. It consists of 3200 fundus photos collected with three distinct fundus cameras, with 46 conditions annotated by two senior retinal experts. This project will enable the development of generalised models for automated classification of different ocular diseases using retinal fundus images.

Details of the dataset is provided in Dataset URL.

Dataset:



(a) (b)

Figure 2. Examples of fundus photographs for (a) healthy eye (b) eye with multiple retinal diseases.

Dataset URL: https://riadd.grand-challenge.org/Data/

Task: Develop a multi-disease detection classifier using neural networks and deep learning algorithms that obtain a performance on the test set with accuracy similar to the grand-challenge submissions leaderboard or come up with novel models with an optimized architecture.

Relevant Papers

[1]. Pachade, S., Porwal, P., Thulkar, D., Kokare, M., Deshmukh, G., Sahasrabuddhe, V. & Mériaudeau, F. (2021). Retinal fundus multi-disease image dataset (rfmid): A dataset for multi-disease detection research. *Data*, *6*(2), 14. https://doi.org/10.3390/data6020014

[2]. Li, N., Li, T., Hu, C., Wang, K., & Kang, H. (2021). A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In *Benchmarking, Measuring, and Optimizing: Third BenchCouncil International Symposium, Bench 2020, Virtual Event, November 15–16, 2020, Revised Selected Papers 3* (pp. 177-193). Springer International Publishing.
https://link.springer.com/chapter/10.1007/978-3-030-71058-3_11

Project ID: 054

Project Title: Diagnosis of Proliferative Diabetic Retinopathy (PDR) and non-PDR using deep learning approach

Area of Research: Computer Vision

Problem Statement: Diabetic retinopathy (DR) is one of the leading causes of blindness, affecting more than 78 percent of people with a 15-year or longer history of diabetes. DR frequently results in subtle alterations in the structure of the vasculature, as well as anomalies. Visual inspection of retinal fundus images for the presence of retinal lesions is normally used to diagnose DR. The discovery of these lesions is crucial in determining whether or not someone has developed DR. There have been some studies that have used fundus images to diagnose DR. OCT angiography (OCTA) is becoming more popular because it can visualise the retinal and choroidal vasculature at a microvascular level in detail; however, the ultra-wide optical coherence tomography angiography imaging (UW-OCTA) modality revealed a higher burden of pathology in the retinal periphery that was not captured by traditional OCTA. The ultra-wide OCTA (UW-OCTA) is a useful imaging modality for helping ophthalmologists diagnose PDR because it can identify DR neovascularization changes non-invasively. However, no works using UW-OCTA that can do automatic DR analysis are currently available. To achieve automatic image quality assessment, lesion segmentation, and PDR detection, it is critical to create a flexible and robust model.

In this project, one/all of three tasks are targeted to complete using the standardized UW-OCTA images provided by the Diabetic retinopathy analysis Challenge (DRAC22) , such as, segmentation of diabetic retinopathy lesions, image quality assessment and/or classification of proliferative diabetic retinopathy (PDR) and non-PDR. Different algorithms will be tested on the DRAC22, and their performance will be tested to make a fair comparison with other algorithms. This project will be an important milestone in automatic image quality assessment, lesion segmentation and DR grading.

Dataset:

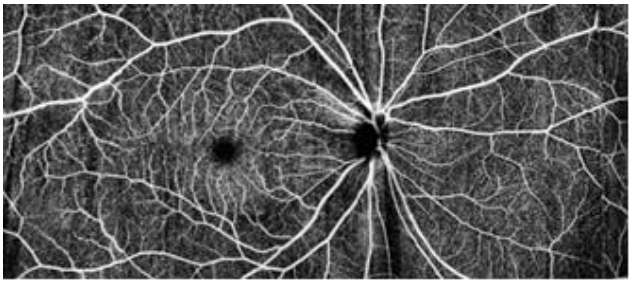


Figure 3. Examples of OCTA retinal images with Proliferative Diabetic Retinopathy (PDR)

Dataset URL: https://drac22.grand-challenge.org/Data/

Task: you can do any of the below tasks:
(a) segmentation of diabetic retinopathy lesions (training set consists of 109 images and test set consists of 65 images.)
(b) image quality assessment (Training set consists of 665 images and test set consists of 438 images)
(c) and/or classification of proliferative diabetic retinopathy (PDR) and non-PDR (Training set consists of 611 images and test set consists of 386 images) .

Relevant Papers

[1]. Qian, B., Chen, H., Wang, X., Guan, Z., Li, T., Jin, Y., ... & Sheng, B. (2024). DRAC 2022: A public benchmark for diabetic retinopathy analysis on ultra-wide optical coherence tomography angiography images. *Patterns*.

https://www.cell.com/patterns/pdf/S2666-3899(24)00020-5.pdf

[2]. Choi, S., Jeoun, B., Anh, J., Jeong, J., Choi, Y., Kwon, D., ... & Shin, S. (2022). A Vision Transformer Based Deep Learning Architecture for Automatic Diagnosis of Diabetic Retinopathy in Optical Coherence Tomography Angiography. In *MICCAI Challenge on Mitosis Domain Generalization* (pp. 135-145). Cham: Springer Nature Switzerland.

https://link.springer.com/chapter/10.1007/978-3-031-33658-4_13