**COMP9444 Project Summary**

Environmental Microorganism Image Analysis Using Deep Learning

Benjamin Wu(z5161230); Jiaqi Wang(z5213190); Raymond Hua(z5214828);

Ruizhi Xu(z5038007); Ziyao Wei(z5306143)

**I.** **Introduction**

The detection and identification of these Environmental microorganisms are crucial in many areas, but the traditional methods involve manual microscopic analysis which is labor-intensive, time-consuming and prone to human error. As a consequence, the dataset available in EMs is highly limited in size, which will be a significant challenge for most machine learning models.

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, and classification.

**II.** **Related Work**

**II.** **I Performance Benchmarking:**

The researchers that released the EMDS-6 dataset published a supplementary paper that provides comparison of 17 CNN models, three VT models, and a hybrid CNN and VT model to test model performance, offering a strong support for related research on feature fusion in order to quickly find suitable models and improve model performance. These experimental results prove that CNN related models such as ‘Xception’ are suitable for these kinds of tasks, and Vision Transformers are highly limited on the small size dataset. (Zhao, Li, Rahaman, Xu, Yang, Sun, Jiang, & Grzegorzek, 2022)

The best accuracy result achieved by this article was 46.03%. Since this article was published in 2022, we aim to investigate whether the rapid development over the past two years and the expansion of pre-trained datasets have led to improvements in CNN pre-trained models for tasks involving extremely small datasets, beyond the initial baseline of 46.03%.

**II.** **II A survey on Image Data Augmentation for Deep Learning**

The study looked at the impact of different data augmentations on applications where there isn’t sufficient data to avoid overfitting. Some of the augmentation techniques discussed include geometric transformations, colour space augmentations, kernel filters and generative adversarial networks. The study goes into details on how these different data augmentation techniques could improve deep learning models as well as limitations of each.

Within the scope of geometric transformations and colour space augmentations, it was found that the best performing augmentation was the cropping method (Taylor, Nitschke, 2018). However, these augmentations were performed on a dataset of approximately 9000 images, significantly more than the 840 images in scope of this report. Further work would need to be done on the applicability of these methods on extremely small datasets which are prevalent in several industries.

The application of generative adversarial networks (GANs) is promising in generating more data for the network to be trained on however it requires a substantial amount of data to train thus is not a practical solution for smaller datasets (Shorten, Khoshgoftaar, 2019).

**III.** **Methods**

To overcome the issue of limited images per mico-organism, numerous noise addition methods, replicating those in the paper, were applied on top of the original images. With 13 different noises applied, including but not limited to gaussian noise, poisson, salt and paper noise, these extra images aim to reduce overfitting by creating better generalisation of the models. On top of these noise added images, filtering techniques such as Weiner, Rank Ordering and Mean Window were applied to help enhance generalisation of model fitting.

**III. I** **Introduction to Autoencoders**

Autoencoders are neural networks that learn efficient data representations for tasks like dimensionality reduction and feature learning. They consist of two parts:

* **Encoder**: Compresses input data into a latent space using convolutional layers and pooling to extract key features and reduce dimensionality.
* **Decoder**: Reconstructs the original data from the latent representation using transposed convolutional layers to upsample the data.

In our project, autoencoders are used for image denoising by minimizing reconstruction loss between noisy and clean images, enhancing image quality.

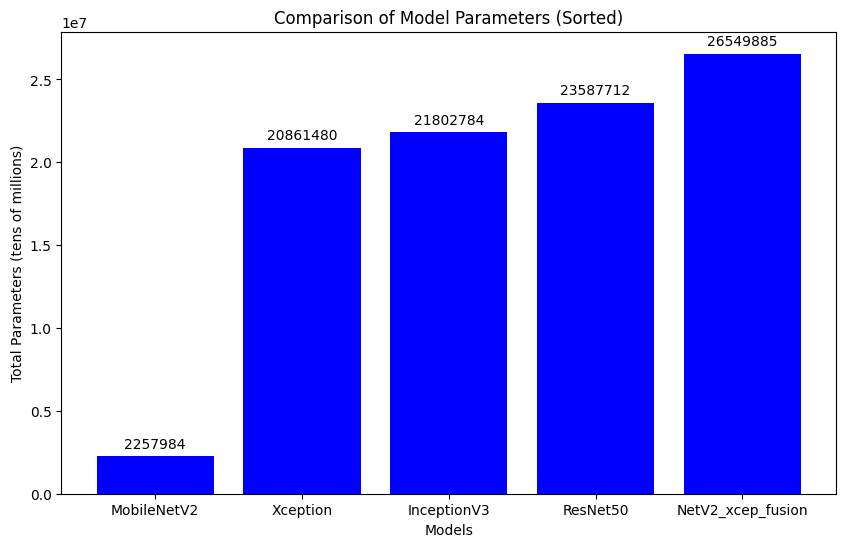
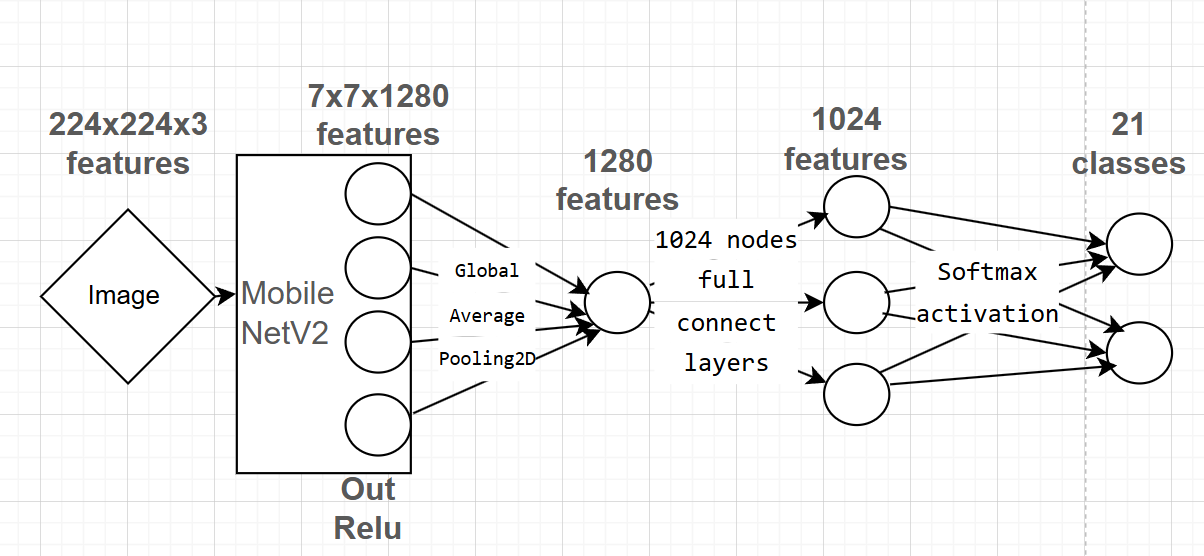
**III.II Data Augmentation:**

We experimented with a few different data augmentation techniques, however they . The best performing augmentation was the cropping method (Taylor, Nitschke, 2018), we utilised our analysis of the dataset to form the parameters of the random cropping. In addition, we applied a random flipping to create more data and variations in the training data set. However, we didn’t apply any more augmentations as this may lead to overfitting of the training data.

**III. III Image Segmentation:**

UNet perform’s best on EMDS-6 dataset among many image segmentation models (Zhao, Li, Rahaman, Xu, Yang, Sun, Jiang, & Grzegorzek, 2022). And considering EMDS-6 is an extremely small dataset, according to the mentor's advice, choosing a pre-trained UNet model instead of training from scratch would achieve performance. Data argumentation tools that used in this section is Albumentations which is a computer vision tool that boosts the performance of deep convolutional neural networks.

**III.IV Classification Methods:**

****

CNN is used in classification to extract features from images since it has been proven that works well in image classification.

We chose the pre-trained model as a feature extractor since this dataset is relatively small, and it is difficult to train a model from scratch fully, we use the per-train models with custom-head work well on our dataset, and it can be achieved with about 80% accuracy on a pre-trained model like MobileNetV2.

Since we found that different model has their advantage on different class’ classifications, therefore the fusion model is created to take advantage of two base pre-train models. The fusion model combined with MobileNetV2 and Xception performs well. We fuse the feature extracted from these two models and then add custom layers, which almost give about 90% accuracy, the best performance classification model we have.

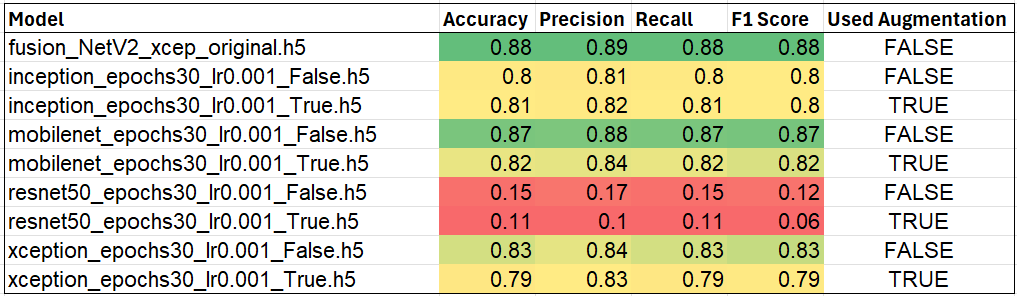
We tried different classification models, for our dataset, the insight is that the pre-train model with too deep a network does not perform well compared to the lightweight model. For example, when using MobileNetV2 as a pre-trained model feature extractor, the performance is 80% accuracy, with about 2,000,000 params. However, the more complex (deep) network ResNet50, has ten times the params compared to MobileNetV2, which has about 20,000,000 params, but ResNet50 cannot even converge in the training, which has terrible results on our dataset.

**IV.** **Experimental Setup**

The dataset (<https://figshare.com/articles/dataset/EMDS6/17125025/1>.) contains original 40 images of 21 different environmental micro-organisms (classes) which are known as the original images. Additionally, there are 40 other images of these environmental micro-organisms which represent the ground-truth, totalling 1680 images. Through data analysis on each of the original images, it was found that the maximum size of the images was 1971 by 1525 px, which disallowed us from adding the performance enhancing technique of padding due to size constraints. Image segmentation was evaluated via the Dice, Jaccardi and Recall measures. The first two measures indicate the similarity between the output and the ground truth segments. On the other hand, recall measures the sensitivity of the true positives to all predicted positives within the segmentations. A learning rate of 0.001 and 30 epochs were used to train the classification model.

**V.** **Results**

Results of the classification models are shown in the table below:



We can clearly see that using the fusion model was the best across all 4 evaluation metrics of accuracy, precision, recall and F1 score with the same number of epochs and learning rate. This was closely followed by the mobilenet model. We believe that the mobilenet is the best given the performance and the low number of parameters, where fusion had significantly more parameters, hence took longer to train.

The above results also show the ineffectiveness of data augmentation on this dataset. All the models which were trained on augmented data performed worse than their counterparts which were trained on unaugmented data. Despite data augmentation providing effective results in other literature, the extremely small dataset resulted in the models overfitting the training data.

Image Segmentation results

**VI.** **Conclusions**

The results demonstrated above show that we were able to effectively use deep learning models to classify the data into the 21 different classes to a relatively high accuracy. For classification, the lightweight model MobileNetV2 obtains the best performance, while the complex model ResNet50 could not converge on this dataset due to the limited size. Furthermore, the MobileNetV2 model provides obvious benefits to the computation time due to the smaller size of the model. Therefore when we do the neural network work, we need to carefully choose the suitable model.

<Describe contribution(s) in the project. What are the key strength(s) of the proposed solution? Describe any limitation(s) of your current study and how it can be improved given more time for the project.>

**VII.** **Reference**

Zhao P, Li C, Rahaman MM, Xu H, Yang H, Sun H, Jiang T and Grzegorzek M (2022) A Comparative Study of Deep Learning Classification Methods on a Small Environmental Microorganism Image Dataset (EMDS-6): From Convolutional Neural Networks to Visual Transformers. Front. Microbiol. 13:792166. doi: 10.3389/fmicb.2022.792166

Shorten, C., Khoshgoftaar, T.M. (2019) A survey on Image Data Augmentation for Deep Learning. *J Big Data* **6**, 60 . doi: 10.1186/s40537-019-0197-0

Taylor, Luke & Nitschke, Geoff. (2018) Improving Deep Learning with Generic Data Augmentation. 1542-1547. doi: 10.1109/SSCI.2018.8628742.