COMP9517 Group Project Report

Team name: Origin

Jinwen Lei

z5435879

z5435879@ad.unsw.edu.au

Yuchen Gao

z5440742

z5440742@ad.unsw.edu.au

Hangwei Liang

z5499015

z5499015@ad.unsw.edu.au

Zhihong Ke

z5388936

z5388936@ad.unsw.edu.au

Tianyi Gao

z5495449

z5495449@ad.unsw.edu.au

***Abstract*—This study presents a comparative analysis of machine models for the automated identification of defects in solar cells from electroluminescence (EL) imagery. This project implement and evaluate four different models—SVM, VGG16, VIT and an innovative CNN. The dataset, comprising 2,624 high-resolution EL images, is preprocessed with techniques like SMOTE to address class imbalance and perform data augmentation to prevent model underfitting. The experimental results reveal that the fine-tuned VGG16 model outperforms others by achieving an accuracy of 77.90% and an F1 score of 79.00% on the test set, closely followed by the innovative CNN model. The study underscores the efficacy of deep learning approaches, particularly CNNs, in capturing intricate image features, leading to high predictive performance in solar cell defect detection. Future work will focus on expanding the dataset, fine-tuning model parameters, and exploring ensemble methods to further improve accuracy and generalizabilit.**

Keywords—solar panel, image recognition, Vgg16, SVM, VIT, CNN

I. INTRODUCTION

Solar panels convert sunlight into electricity through photovoltaic (PV) cells, providing a cost-effective, renewable, clean energy solution. Commercial solar panels use protective strategies such as enclosures to protect the solar panels from rain, wind, snow, and other impacts. However, solar panel quality can still be affected by natural factors or errors in the manufacturing process. Any defects can have a huge impact on solar panel power efficiency, so monitoring solar panel quality is very important. Manually inspecting the condition of solar panels is labor-intensive and ineffective in identifying damage. In this case, it is more advantageous to use automatic recognition, which is faster and more accurate. Photovoltaic modules scanned by electroluminescence imaging technology can well display defects and damage. Defective components will appear darker than normal components, and these images are analyzed through computer vision methods to detect and classify defects.

The ELPV data set [1] provides images of solar cells under EL conditions. Each sample contains three labels: image, loss probability, and category, which can be used to evaluate the quality of solar cells. Additionally, it can be used for image segmentation, such as deep learning and convolutional neural networks, to reveal battery defects and efficiency changes.

II. LITERATURE REVIEW

"A Benchmark for Visual Identification of Defective Solar Cells in Electroluminescence Imagery" (Buerhop-Lutz et al.) [2]

This study analyzes a dataset and utilizes 2624 high resolution solar cell images of that, trying to develop machine learning techniques of automatic defect detection of solar cells. This study depends on defect likelihood to categorize cells and highlights the development methods for predicting power efficiency loss due to defects.

"Automatic classification of defective photovoltaic module cells in electroluminescence images" (Deitsch et al.) [3]

The study uses 1,968 cells high-resolution EL images from same dataset and finds that the CNN model has higher accuracy (88.42%) compared to the SVM (82.44%). This study recommends using SVMs for rapid evaluation and CNNs for more comprehensive analysis. In addition, this research

focuses on the defect detection of SVM and deep CNN in photovoltaic cells.

III. METHODS

A. Pre-processing

In the data preprocessing part, many innovations have been made. First, since the probability in the label has multiple digits (such as 0.33333, 0.66666), for better observation, the probability is rounded and mapped to four probabilities of 0, 1, 2, and 3. Secondly, after printing out the number of training samples for each probability in the data set for analysis, it was found that there is a data imbalance problem (the number of label 0 is much larger than other labels). After splitting 75% of the samples into the training set and 25% of the samples into the validation set. We have performed data balancing on the training set. If we train based on the original unbalanced data, the model is likely to directly identify all samples as 0. This will improve the accuracy, but it is not the goal of training the model. In order to achieve relatively balanced conditions, the SMOTE method is used to balance the data set, which mainly reduces the number of majority samples (label 0) by randomly deleting samples. For minority samples (label 1, label 2, label 3), the SMOTE [4] method first randomly selects a minority sample, and after finding the nearest neighbor sample of the sample, uses the feature difference between the two samples as interpolation and adds random numbers to merge into new samples to increase the sample size. In this way, SMOTE can generate new, non-duplicate samples, which helps avoid overfitting and improves the model's predictive ability for minority classes. The model was also tested on the balanced data set and the original imbalanced data set to compare whether data imbalance would have an impact on model training. Third, the data set itself has a small number of samples, which may lead to underfitting. In order to solve this problem, eight data enhancement methods (noise, cropping, flipping, rotation, scaling, adjusting brightness, adjusting contrast, adjusting saturation) were defined.

After increasing the number of training sets through these eight methods, the number of training sets increased from 2520 The number of images increased to 5040, allowing the model to have more training data, effectively preventing model underfitting and improving the accuracy of model predictions. At the same time, the image size in the data set was originally 300x300, which took too much time to train the model. In order to speed up the training, the image size was adjusted to 64x64 through the resize method of the cv library. After adjusting the image size, the training speed was significantly accelerated. Finally, in order to explore whether the "single crystal" or "poly crystal" of the battery components will have an impact on the training model, the "single crystal" and "poly crystal" data sets were divided, and then the divided data sets were analyzed again. Equalize and enhance treatments to compare whether single or polycrystals affect training. After data processing, a total of six different sets of training sets and validation sets were obtained (data sets before and after balancing, each part is divided into all data, poly data and mono data). With these six different sets of data differences as data sets, the impact of the data on model training can be well analyzed through evaluation methods such as Accuracy, F1 Score, and confusion matrix.

B. Support vector machine(SVM)

SVM is a classification algorithm by finding a classification plane and separating the data on both sides of the plane to achieve the purpose of classification. SVM is a highly efficient machine learning model that has significant application areas. The areas that SVM can be applied to include facial recognition, character recognition, Pedestrian detection, and text classification. In machine learning, SVM is a Supervised learning model which is used for pattern recognition, classification (outlier detection) and regression analysis [5].

This article combines SVM with random forest and set RBF as the kernel.

To begin with, the model uses a reshape function to guarantee that all data is converted smoothly into one-dimensional arrays. After reshape operation, random forest with 100 trees and random state of 42 is applied to implement feature selection. This is because the model can achieve the highest accuracy at this value. Train SVM model after the features is extracted, the result displays that the ability of SVM model has a slight improvement.

On the one hand, this method reduces the risk of overfitting, on the other hand, it improves predictive ability of model on unknown data.

C. *Vgg16*

Vgg16 is a deep-learning model that is widely used in computer vision to process grayscale images, extract features for classification and implement classification or regression tasks[6].

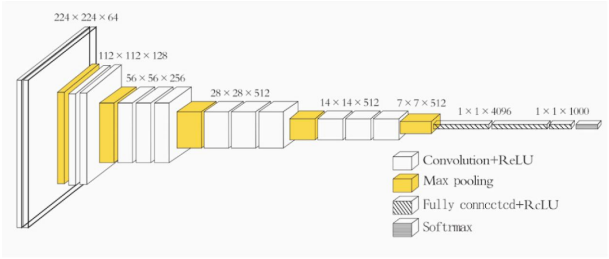

Fig. 1. VGG16 Concept Diagram [7]

In this article, to utilize a pre-trained VGG16 neutral network to classify images, model adaptation for specific datasets and evaluation of its performance.

At the first, the image set are sized to meet the input requirements (64*64*3), and the grayscale images are converted to RGB format to meet the VGG16 model requirement. Then, using pre-trained VGG16 network, it is enhanced with a custom layer to tailor it for specific classification tasks. To train different datasets in the model, capturing performance history. Finally, the results are visualized according to accuracy and loss curves, along with heatmaps of confusion matrices, providing in-depth understanding of the model abilities and areas for improvement in image classification.

In essence, this approach illustrates the practical application of a pre-trained deep learning model in custom image classification tasks. It underscores the significance of preprocessing in aligning datasets with model requirements and the value of performance evaluation in understanding and improving model effectiveness. The visualization of results through accuracy and loss curves, along with confusion matrix heatmaps, offers comprehensive insights into the model's classification capabilities, paving the way for further optimizations and adaptations in the realm of image processing and computer vision.

D. *CNN*

Convolutional neural networks (CNN) are a deep learning architecture widely used in the fields of computer vision and image processing. They excel in fields such as image recognition, object detection, and video analysis. The key feature of CNN is its ability to learn features directly from image data without the need for manual feature extraction.

The model is innovative based on a traditional CNN architecture, incorporating some deep learning techniques such as Residual Connections, Depthwise Separable Convolutions, and Spatial Pyramid Pooling. These technologies can improve the performance of the model while reducing the demand for computing resources.

- Depthwise Separated Convolutions [8]: The innovative model uses deep separable convolutional layers, which are divided into two parts: deep convolution and point by point convolution. This structure has higher computational efficiency and fewer parameters than standard convolution, helps reduce overfitting, and is suitable for mobile and embedded devices.

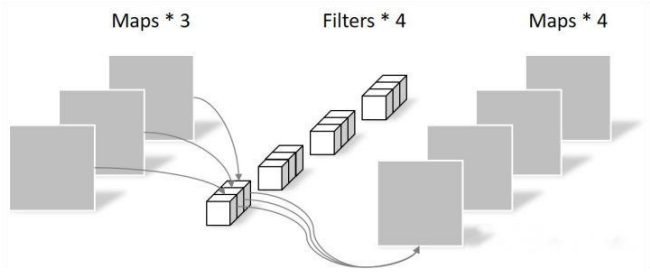

Fig. 2. Point by Point Convolution [8]

- Residual Connections: The model includes residual connections, which helps solve the gradient vanishing problem in deeper networks. Residual connections allow gradients to flow directly through the network, improving the stability and efficiency of training.
- Global Average Pooling: Use global average pooling after the final convolutional layer instead of the traditional fully connected layer. This reduces the number of parameters in the model, helps reduce overfitting, and reduces computational complexity.
- Multi-layer Dropout: The innovative model uses Dropout layers before and after the fully connected layer, which helps to provide additional regularization and prevent overfitting during the training process.

E. *Vision Transformer (ViT) Model – 8 patch*

In the field of computer vision, traditional Convolutional Neural Networks (CNNs) have achieved significant results. However, with the development of deep learning technology, Transformer models have been introduced into image recognition tasks due to their superior performance in handling sequential data. The Vision Transformer (ViT) model is an emerging neural network architecture that processes images as a sequence of data by dividing them into multiple small patches and mapping these patches into embedded vectors through linear transformations. Key components of the ViT model include linearly projected image patches, position encoding, and the Transformer encoder.

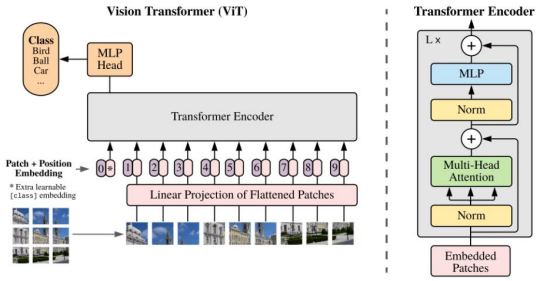

Fig. 3. Architecture of a ViT for Image Classification [9]

In this study, we have implemented a ViT model based on PyTorch for image classification tasks. By preprocessing input images to convert them into a single-channel format acceptable to the model and adding positional information, we have successfully trained the model to recognize images of different categories. Within the model's structure, we utilized multi-head self-attention mechanisms to capture the complex features within images, while the multi-layer perceptron (MLP) head classified these features.

The training of the model utilized cross-entropy loss function and the AdamW optimizer, along with a learning rate scheduler to optimize the training process. In each training epoch, we evaluated the model's performance using multiple indicators such as accuracy and recall measuring the model's classification effect. In addition, we visualized the loss and accuracy during the training and validation processes to more intuitively observe the model's learning progress.

F. *Vision Transformer (ViT) Model – 16 patches*

ViT_B_16 (Vision Transformer Base 16) [8] is a deep learning model with applications in computer vision. It is a variant of the Vision Transformer (ViT) family of models, first proposed by Google in 2020. This model marks the successful application of the Transformer architecture in the field of natural language processing to the task of image recognition.

Basic structure: The ViT_B_16 model splits the input image into patches of size 16x16 pixels. These patches are linearly embedded into a higher dimensional space, similar to word embedding in natural language processing. These embeddings are then fed into a standard Transformer structure. The Transformer structure relies heavily on the Self-Attention mechanism to efficiently handle the relationships between each element in the input sequence. This is very useful in image processing as it allows the model to capture complex relationships between different parts of an image. ViT_B_16 demonstrates excellent performance on a wide range of image recognition tasks, including image classification, target detection, and image segmentation. It is particularly good at handling large-scale datasets such as ImageNet.ViT shows better performance and higher data efficiency compared to traditional convolutional neural networks (CNNs). It is able to process global information in images more efficiently, which is especially important when dealing with complex scenes.

IV. EXPERIMENTAL RESULTS

A. *Accuracy*

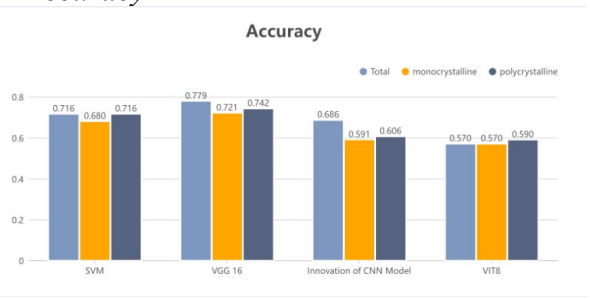

Fig. 4. Accuracy of four Model

B. *F1*

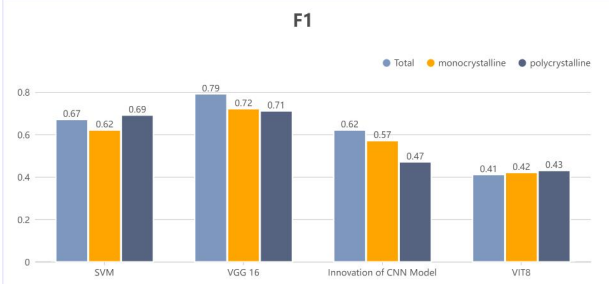

Fig. 5. F1 of four Model

C. confusion matrix

- SVM

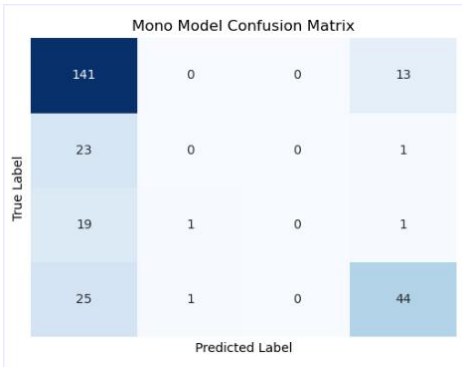

Fig. 6. Mono Model Confusion Matrix for SVM

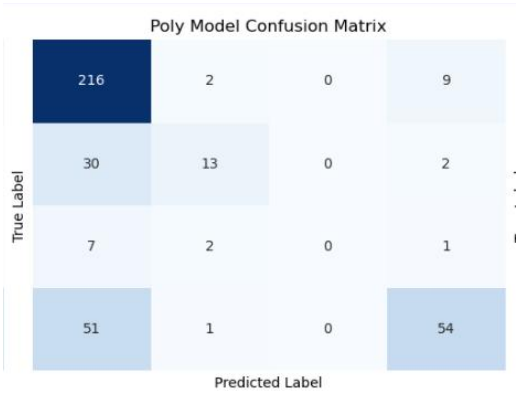

Fig. 7. Poly Model confusion matrix for SVM

Fig. 8. Total Model Confusion Matrix for SVM

- VGG16

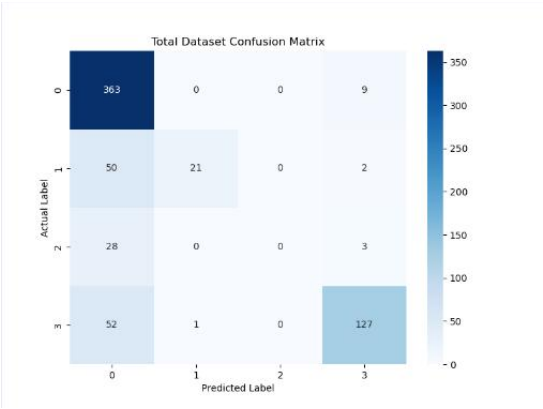

Fig. 9. Mono Model Confusion Matrix for VGG16

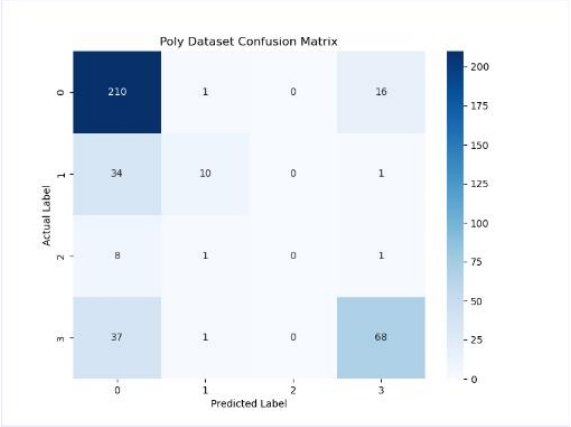

Fig. 10. Poly Model Confusion Matrix for VGG16

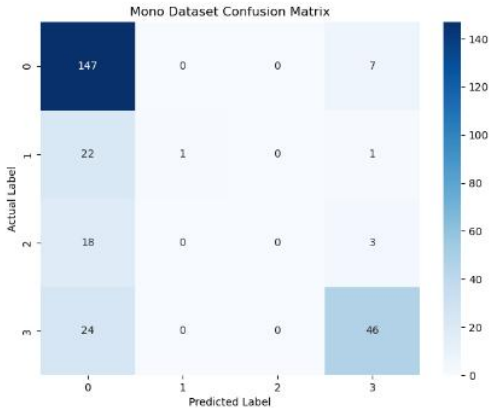

Fig. 11. Total Model Confusion Matrix for VGG16

- Innovation of CNN

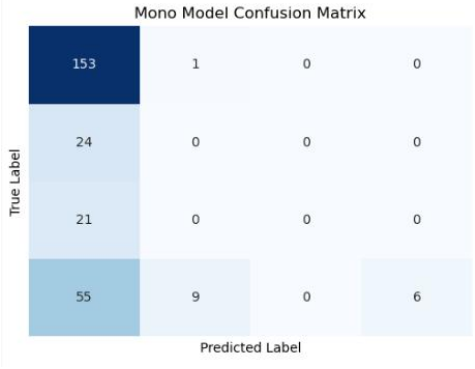

Fig. 12. Mono Model Confusion Matrix for CNN

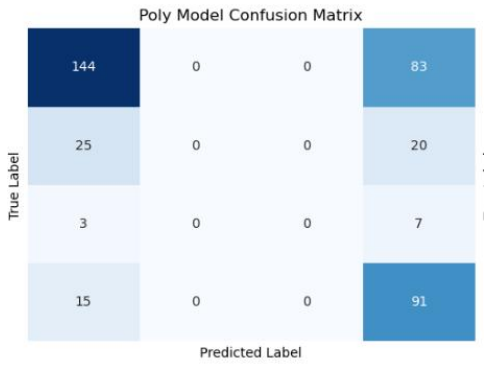

Fig. 13. Poly Model Confusion Matrix for VGG16

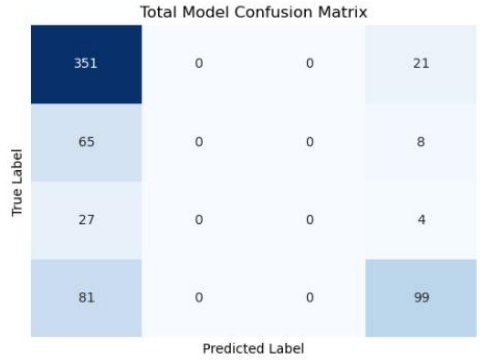

Fig. 14. Total Model Confusion Matrix for VGG16

- VIT

Mono

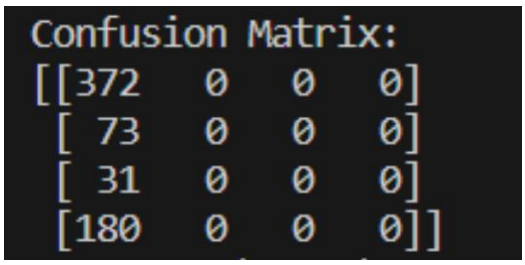

Fig. 15. Mono Model Confusion Matrix for VIT

Poly

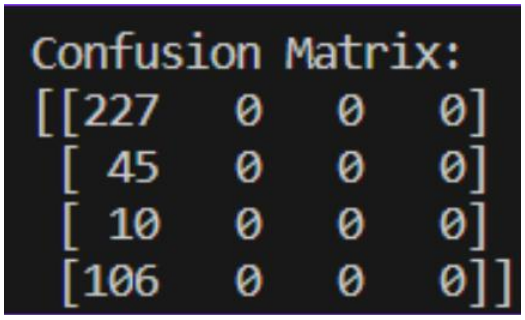

Fig. 16. Poly Model Confusion Matrix for VGG16

Total

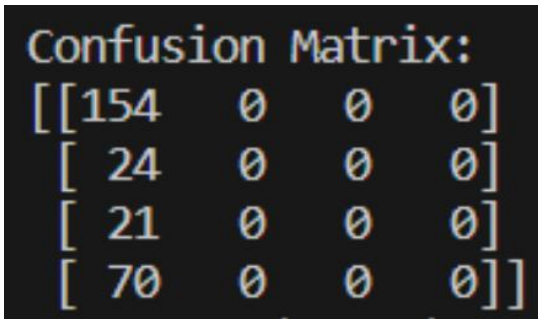

Fig. 17. Total Model Confusion Matrix for VGG16

V. DISSCUSSION

A. Model Analysis

- SVM

In this project, random forest is implemented into SVM model. SVM is a classification model which is advantaged in small dataset and high dimensional space. It is able to prevent models from overfitting while another model is possible to meet this situation. Set RFB as the Kernal, the RBF kernel mainly has two parameters: C and gamma. The features of solar panel images have no-liner relation with labels, RBF Kernal has advantaged performance in this relation and achieve accurate classification. Set the tree of random forest as 100 and random state as 42. This is because 100 trees can balance performance and computing cost, and random state can ensure consistency of results and facilitate reproduction and comparison. These modifications can improve the accuracy and generalization ability of SVM models, also identify the most informative features to improve the data processing capabilities for handling non-linear and complex patterns. In addition, combining random forest with SVM can reduce the risk of overfitting. According to the confusion matrix, classification on class one can achieve the highest accuracy, while SVM performs the lowest accuracy on class 3. Class 2, 3, 4 have a higher possibility to be recognized as class 1. This is possibly because of the mislabeling of labels and insufficient extraction of features. In conclusion, correcting the labels, applying data augmentation strategies and cross-validation can be applied to achieve more accurately evaluate model performance.

- VGG16

In this project, the pre-trained VGG16 model demonstrates robust application in image classification with and accuracy rate of more than 70%, showing advanced features learning from

ImageNet and adaptability to huge datasets. However, it still faces some challenges, such as overfitting and computational demands. But these can be prevented by taking some measures, such as adding early stopping and dropout.

According to three confusion matrices, it is obvious to find that the VGG16 model performs best for class 0. However, it is struggling with class 1 and 2 for all data, which means there exists an issue with training or class imbalance. For this issue, adding more training data, adjusting the model's settings can be used to reduce it.

Overall, the VGG16 model pre-trained on ImageNet, is adeptly customized and trained on various datasets for image classification, showcasing its efficiency and adaptability through detailed performance evaluations and visualizations.

- Innovation of CNN Model

For this model, the accuracy of all test cell images together (monocrystalline + polycrystalline) is approximately 68.96%, which is the highest among the three datasets, indicating that the model has good predictive ability on the comprehensive dataset. The accuracy of the monocrystalline and polycrystalline datasets is slightly lower than that of the total dataset.

From the confusion matrix of the three datasets, the model tends to predict the majority of samples as class 0, which may reflect imbalanced categories in the dataset or excessive learning of features from class 0. On single crystal and polycrystalline datasets, the model is almost unable to accurately identify categories 1 and 2, indicating that the model may not have captured the key features that distinguish these categories. On the overall dataset, the model can identify categories 0 and 3 with high accuracy.

Overall, the performance of the Innovation of CNN model overall dataset is superior to that of single crystal and polycrystalline datasets, possibly because the overall dataset provides more features and sample diversity, which helps model learning and generalization. Due to class imbalance, the training and evaluation of the model can also consider using data resampling or class weighting to improve the recognition ability of minority classes.

- VIT8

The results provided indicate that there is a variation in performance when the Vision Transformer (ViT) model is trained and evaluated on different datasets. The accuracy consistently remains at a value of 0.57, which may be due to issues with the model itself or with the training data. However, the classification report shows high precision for certain categories, while others are not recognized at all, which may be attributed to class imbalance or the model's inability to generalize across all categories. Although the accuracy across different datasets is moderate, the F1 scores for specific categories are notably low, indicating poor recall. Overall, despite the model demonstrating good precision in classifying certain types of crystalline structures, there is significant room for improvement in enhancing the model's robustness and addressing class imbalance.

The complete absence of classifications for classes beyond the first one, as observed in the confusion matrix, is indicative of a model with a significant bias or a severe class imbalance issue. Such an outcome may stem from inappropriate model complexity, inadequate feature representation, or underfitting due to limited and imbalanced data. It may also suggest that the model has failed to capture the discriminative features essential for distinguishing between the various classes.

- VIT16

In training, the VIT model is very slow, which may be caused by the complex model architecture. The original batch size of the model was 32, and due to the performance of the device (not enough gpu memory), it was not possible to run the model at first, and after modifying the batch size to 8, the model was finally able to run slowly. The original model used the SGD optimizer, which is a good optimizer but requires keeping an eye on the model to manually adjust the learning rate and so on. I replaced the optimizer with the Adam optimizer, which automatically adjusts the learning rate and is a convenient choice for faster convergence in the early stages of training. The model initially chose Focal Loss for the loss function, and since we have already processed the data equalization in the data preprocessing stage, I replaced the loss function with cross-entropy loss, which is a very common loss function for multi-classification tasks. After the modification, I trained the model for a total of 30 epochs, which took about three hours or so, and the model achieved roughly 68% accuracy on the dataset,

which is actually not a very good training result. In response to that low training result, I guess there may be the following reasons, the VIT model itself needs a lot of data to support it, and that dataset is too small, so the model is not good. Secondly, the dataset seems to have some mislabeling (labeling that can easily confuse the model), there are images that seem to be a bit damaged to the naked eye, but the LABEL labeling is 0.0. It is possible that the model is not as good as it should be based on the above reasons.

B. Summary Analysis

VGG 16 has the highest accuracy and F1 score on all datasets, indicating its strong ability in deep learning structures for image classification tasks. Secondly, on the mono and poly datasets, the results of SVM are relatively high, indicating that SVM has good recognition ability for specific feature spaces. Afterwards, the CNN model performed moderately on the comprehensive dataset, but there was a significant decrease in performance on the mono dataset. This may be due to the CNN model failing to capture key features of mono data. Finally, VIT performed the worst on all datasets, which may be due to the structure of the VIT model not being suitable for the this data features, or due to insufficient training.

From the perspective of confusion matrix, all models show a tendency to over recognize class 0, which may be due to the large number of class 0 samples in the dataset, resulting in the model learning leaning towards this majority class. VGG16 has the best performance in identifying category 3, indicating that its deep learning architecture can capture more complex features and help distinguish samples from different categories. Overall, VGG16 exhibits better adaptability and performance when generalized to different datasets (single crystal, polycrystalline, and population). Improving the recognition ability of a few categories is a common challenge for all models. This issue may need to be addressed through resampling, increasing sample weights, or improving feature engineering.

Overview, the VGG 16 model is the best choice because its deep network structure is suitable for complex image classification tasks and can handle different datasets well. Secondly, SVM still performs well in processing relatively simple image data, but may not be suitable for complex image features. As for CNN model innovation and VIT8, they may need to be further adjusted and optimized to improve performance, such as adjusting network structure, hyperparameter adjustment, and ensuring sufficient training data.

Afterthat, four models are implemented to compare the performance on image recognition of solar panel. Universually, models can reach over 70% of accuracy on poly images but commonly lower than 70% on mono images. This is possibly because that poly images have more apparent features such as color, texture and reflective properties. In addition, label errors, data bias and training data imbalance is also probably cause this result.

C. Compare with the Creator

The creators of the ELPV dataset utilized a deep regression network (based on a pre-trained VGG network) that was trained on 2,426 high-resolution electroluminescence (EL) images to automatically identify the probability of defects in solar cells. This network achieved an accuracy of 88.42% and an F1 score of 88.39% on the test set, along with a ROC AUC value of 94.7%. This represents a very high level of performance, particularly in the automatic detection of defects in solar cells.

Using the best performing VGG16 as a comparison, VGG16 performed very closely on the overall dataset compared to the model in PDF, but slightly decreased on the single crystal and polycrystalline datasets. This may be because the model may not have fully reached the training level described in the PDF, or the data distribution may differ during training and testing.

VI. CONCLUSION

According to the analysis in the results section, the VGG 16 model is the best choice because its deep network structure is suitable for complex image classification tasks and can handle different datasets well. However,VGG16 Is a deep network with a large number of parameters, which can lead to overfitting, especially with low amounts of data. Also, it means higher computational costs and longer training time.

Secondly, SVM still performs well in processing relatively simple image data but may not be suitable for complex image features.In summary, while SVMs are powerful tools for image classification tasks with simpler datasets, their efficacy diminishes as the complexity of image data increases. The challenges posed by complex features, high dimensionality, scalability,

and dependency on feature engineering make SVMs less suitable for advanced image processing tasks compared to more sophisticated algorithms like convolutional neural networks (CNNs).

Aftertat, the innovative CNN model is an innovation based on traditional CNN models, combining some deep learning technologies such as residual concatenation, deep separable convolution, and spatial pyramid pooling. These technologies can improve the performance of the model while reducing the demand for computing resources. However, overall, the performance of CNN models is poor, especially on polycrystalline datasets, which may require improvements to the model architecture or training methods.

As for the VIT model, this model is a milestone of the transformer architecture in the image field, but its performance on this data set is very poor. The reason for this may be that the model requires a large amount of training data to support. Because the data set is too small, the VIT model lacks inductive bias, so its performance is very poor.

The accuracy of almost all models is not very good. After analysis, the possible reasons are found. After looking at the data set, it is found that some pictures (cell 811) with poor performance and dark appearance have a damage probability of 0%. These pictures can easily make the model Confusion occurs and useful information cannot be learned well.

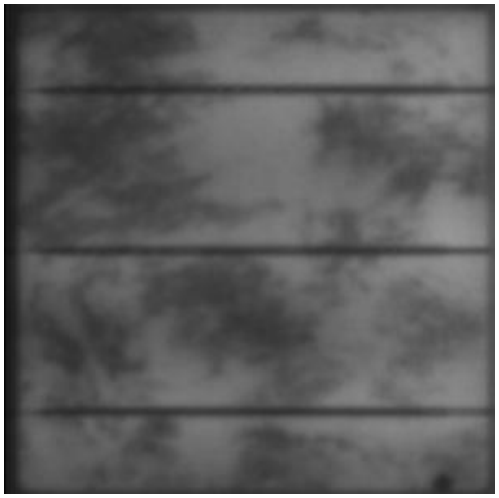

Fig. 18. Poor Performing Images - Cell0811 [1]

Second, the dataset size is very small, and too little training data may result in model underfitting, weak generalization ability, and reliance on specific data. Due to limited device performance, almost all models only ran for 30 epochs, which may also be one of the reasons affecting the final performance of the models.

To solve these problems, images that are mislabeled or easily confused by the model can be corrected by relabeling or deleting easily confused samples. The data set itself can then be increased in size by collecting more data. Using cross-validation to evaluate a model's ability to efficiently utilize data with small data sets is effective. In addition, you can observe whether there will be better recognition results by thresholding the image after observing the color histogram. After changing to better equipment or renting a cloud server, you can try to use more deep models and run more epochs to improve the accuracy of image recognition.

VII. REFERENCES

[1] ELPV Dataset. A Benchmark for Visual Identification of Defective Solar Cells in Electro luminescence Imagery. https://github.com/zae-bayern/elpv-dataset

[2] C. Buerhop et al. A Benchmark for Visual Identification of Defective Solar Cells in Electro luminescence Imagery. European PV Solar Energy Conference and Exhibition (EU PVSEC), 2018. http://dx.doi.org/10.4229/35thEUPVSEC20182018-5CV.3.15

[3] S. Deitsch et al. Automatic Classification of Defective Photovoltaic Module Cells in Electro luminescence Images. Solar Energy, vol. 185, June 2019, pp. 455-468. https://doi.org/10.1016/j.solener.2019.02.067

[4] N. V. Chawla, K. W. Bowyer, L. O.Hall, W. P. Kegelmeyer, ″SMOTE: synthetic minority over-sampling technique,″ Journal of artificial intelligence research, 321-357, 2002.

[5] Karaağaç MO, Ergün A, Ağbulut U, Gürel AG, Ceylan I, ′Experimental analysis of CPV/T solar dryer with nano-enhanced PCM and prediction of drying parameters using ANN and SVM algorithms′, Solar Energy, vol.218, accessed 14 November 2023, < https://doi.org/10.1016/j.solener.2021.02.028>.

[6] S.Sarker, S,Tushar,H.Chen,"High accuracy keyway angle identification using VGG16-based learning method",Journal of Manufacturing Processes,vol.98,pp.223-233,2023,https://doi.org/10.1016/j.jmapro.2023.04.019

[7] Y. Chen, Y. Chen, S. Fu, et.al, ″VGG16-based intelligent image analysis in the pathological diagnosis of IgA nephropathy″, Journal of Radiation Research and Applied Sciences, vol.16, 2023, https://doi.org/10.1016/j.jrras.2023.100626

[8] Hong, G., Chen, X., Chen, J. et al. A multi-scale gated multi-head attention depthwise separable CNN model for recognizing COVID-19. Sci Rep 11, 18048 (2021). https://doi.org/10.1038/s41598-021-97428-8

[9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, ... N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929, 2020. [Online]. Available: https://arxiv.org/abs/2010.11929