ECE 157A
Lab Report 3

The objective of this lab was to develop and train machine learning models to classify images from three distinct datasets—semiconductor wafer maps, pigmented skin lesions, and natural object images—each requiring unique preprocessing and modeling approaches. For wafer maps, custom models and VGG16 were built to detect defect patterns; for skin lesions, both custom models and a fine-tuned VGG16 model were used to navigate the high variability in medical images; and for the CalTech-256 natural images dataset, similar methods were applied to handle diverse and complex visual data. The work involved detailed preprocessing, model architecture design, optimization parameter selection, and the evaluation of model performance based on accuracy metrics, offering practical insights into the application of both bespoke and transfer learning techniques in image classification.

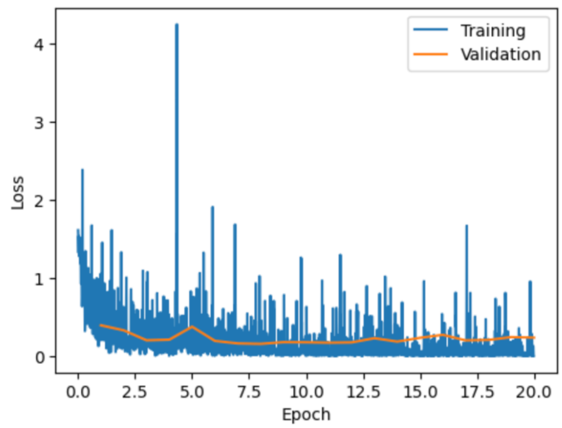
Task 1.1:

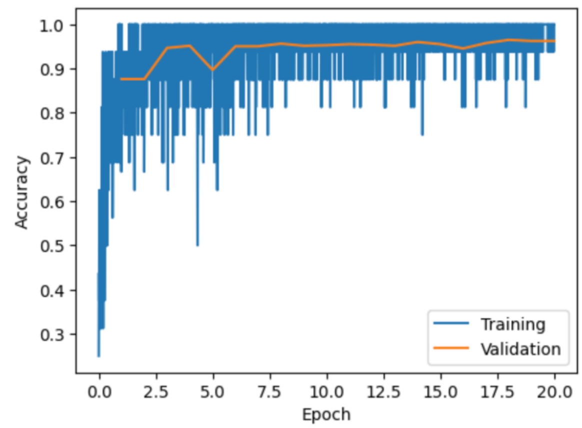
- **Data Preprocessing and Split Ratio:** The preprocessing involves downloading and unzipping the dataset, loading it into a Pandas DataFrame, and resizing wafer maps to a uniform size of 64 × 64 pixels. The dataset is then split into training and validation sets with a 70%-30% ratio, ensuring a balanced representation of different failure types in both sets.

- **Model Structure:** The WaferNetwork model is a convolutional neural network comprising two convolutional layers with 32 and 64 filters, respectively, followed by max pooling layers. It also includes two fully connected layers, with dropout applied for regularization. The ReLU activation function is used throughout the network. This structure is designed to effectively learn from the resized wafer maps.

- **Optimization Parameters:** The model is optimized using Stochastic Gradient Descent (SGD) with a learning rate of 0.1. Cross-entropy loss is employed as the loss function. The training is conducted over 20 epochs, with performance metrics like loss and accuracy being tracked and visualized to assess the model's learning progress and effectiveness.

- **Visualization of the model's loss and accuracy over the training process:**





- **Model Accuracies:**

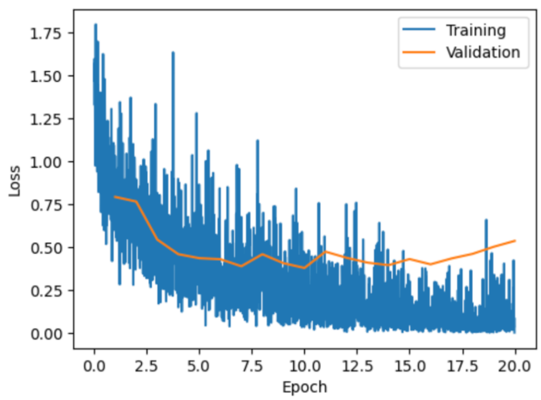
  1) Training: around 90%
  2) Validation: 95.6%
  3) Test data: 89.2 %

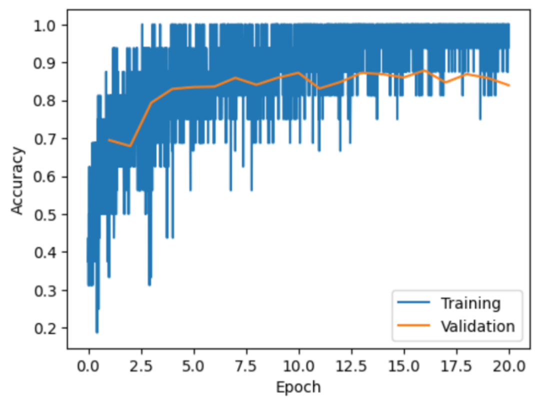
Task 1.2:

- **Data Preprocessing and Split Ratio:** In this section, the code adapts a pre-trained VGG16 model for the wafer defect classification task. The data preprocessing involves resizing the wafer images to 224x224 pixels to match the input size expected by VGG16. The dataset is then re-split into training and validation sets using a manual seed for reproducibility, maintaining the previous split ratio of 70% for training and 30% for validation. The wafer dataset is tailored to the model by modifying the last fully connected layer of VGG16 to output 5 classes, corresponding to different wafer defect types. Additionally, the feature extraction layers of VGG16 are frozen during training to retain the knowledge from ImageNet, while only the last layer's weights are fine-tuned to the wafer dataset.

- **Optimization Parameters:** The optimization for fine-tuning the VGG16 model utilizes Stochastic Gradient Descent (SGD) with a learning rate of 0.001 and momentum of 0.9. These parameters are chosen to fine-tune the pre-trained network's last layer on the new dataset, with the aim to transfer the learned features from the ImageNet dataset to the wafer defect classification task. The model is trained for a number of epochs defined earlier in the code (not specified in the given excerpt), and the training process involves tracking loss and accuracy metrics to monitor and visualize the model's performance over time.

- **Visualization of the model's loss and accuracy over the training process:**





- **Model Accuracies:**
  1) Training: around 90%
  2) Validation: 87.2%
  3) Test data: 70.7 %

Task 1.3 :

- **Key Characteristics of the Semiconductor Wafer Map Dataset vs. the Models:**
  The specific challenges of semiconductor wafer testing datasets include the high variance in defect patterns and the subtle distinctions between different defect types. A custom neural network may be more suitable for this domain if it's specifically designed to recognize these patterns, as it can be tailored to the unique characteristics of wafer maps. On the other hand, a pre-trained model like VGG16 offers the advantage of transfer learning, where knowledge from a vast and diverse dataset (like ImageNet) can be leveraged. The complexity inherent in the wafer maps, which includes noise and irregular patterns, could be a challenge for VGG16 if these features are significantly different from the natural images it was trained on. The custom model could potentially be simpler and more focused on the specific domain, reducing computational load and avoiding the need to adapt to features irrelevant for wafer maps.

- **Selecting a Model Considering the Constraints:**
  Considering data availability and characteristics, a custom neural network might be preferred for training a model on wafer map patterns. This preference arises because custom models can be designed to focus on the specific features of wafer maps, which

may not be well-represented in the pre-trained layers of foundation models like VGG16. Furthermore, a custom model can be built to be less complex and more interpretable, which is important in a domain where understanding the model's decision-making process can be as crucial as its performance.

- **Achieving 99% Accuracy without Increasing Data Size:**
  To achieve 99% accuracy without increasing the data size, the focus would be on improving the model's ability to generalize from the available data. This could involve techniques such as data augmentation to artificially increase the dataset's diversity, regularization methods to prevent overfitting, and hyperparameter tuning for optimal performance. In this scenario, a feature-based model or a custom neural network might be more amenable to such fine-tuning, as they can be adjusted more readily without the constraints of a pre-trained model's architecture. The training process would need to be modified to include a more rigorous validation scheme, possibly using techniques like k-fold cross-validation to ensure the model's performance is robust across different subsets of the data.

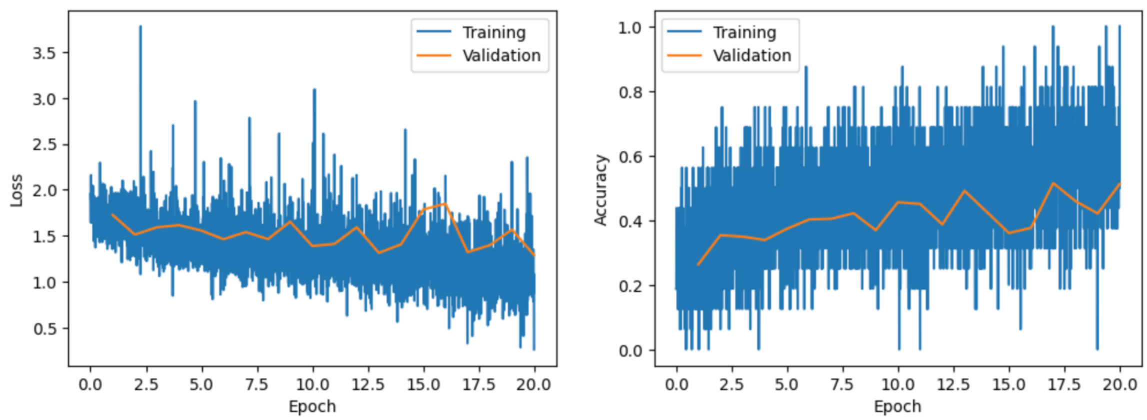
Task 2.1 :

- **Data Preprocessing Steps and Data Split Ratio:**
  For the HAM10000 dataset, the preprocessing steps include resizing images to 150x200 pixels and converting them to tensors suitable for model input. The data split ratio is maintained at 70% for training and 30% for validation, similar to the previous dataset. This split ratio ensures that a sufficient amount of data is used for training the models while still providing a robust validation set to evaluate the models' performance. This careful partitioning of the data is crucial for the development of a reliable classifier, allowing the model to learn a diverse array of features from the training set and then validating its ability to generalize to unseen data.

- **Model Structure:**
  The custom neural network model, HAM10000Network, designed for the HAM10000 dataset consists of a series of convolutional layers followed by max pooling layers and dropout for regularization. Specifically, it includes four convolutional layers with progressively increasing filters (32, 64, 128, 128) and two fully connected layers, where the final layer outputs the probability distribution over seven classes. The use of dropout helps mitigate overfitting, allowing the model to generalize better to new data. The input size for the fully connected layer is adjusted based on the transformed image size and the effect of the convolutional and pooling layers.

- **Optimization Parameters:**
  The optimizer used for the custom model is Stochastic Gradient Descent (SGD) with a learning rate of 0.1. SGD is a well-established optimization algorithm that can navigate the complex loss landscapes of neural networks effectively. The learning rate of 0.1 is chosen to ensure a balance between the speed of convergence and the risk of overshooting minima in the loss function. The model is trained over a fixed number of epochs, with an early stopping mechanism based on validation loss to prevent overfitting and to save the best performing model.

- **Visualization of the model's loss and accuracy over the training process:**



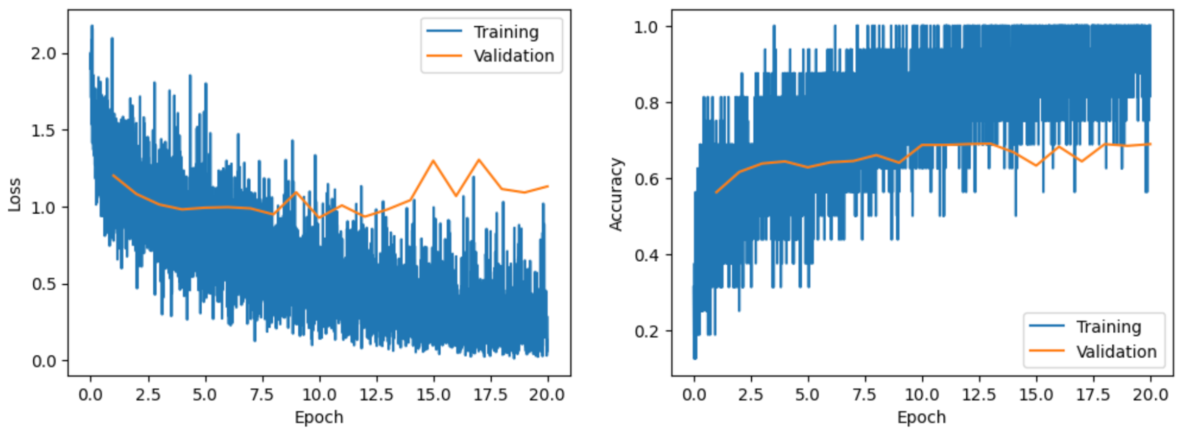
- **Model Accuracies:**
  1) Training: around 70%
  2) Validation: 51.2%
  3) Test data: 47.4 %

Task 2.2 :

- **Data Preprocessing Steps and Data Split Ratio:**
  The data preprocessing for the HAM10000 dataset includes a custom transformation that pads images to a square shape, followed by conversion to a tensor, resizing to 224x224 pixels, and normalization with specified mean and standard deviation values. These steps prepare the data for processing by the VGG16 model, which is designed to work with square input images of size 224x224. The dataset is split into training and validation sets with an 80-20 ratio. The larger portion for training allows the model to learn from a more comprehensive range of data, while the validation set is sufficient to evaluate the model's performance and generalization capabilities.

- **Optimization Parameters:**
  The VGG16 model is optimized using Stochastic Gradient Descent (SGD) with a learning rate of 0.001 and a momentum of 0.9. This choice of optimizer and parameters is intended to finely tune the pre-trained network to the specific task of skin lesion classification. The relatively low learning rate is chosen considering that VGG16 is already pre-trained on ImageNet, and drastic updates could potentially degrade the learned features. Momentum is employed to accelerate convergence and to help navigate the complex optimization landscape of deep networks. The model is trained over a series of epochs (the exact number was specified earlier in the code but not in the provided excerpt), and the training process is carefully monitored to save the best-performing model based on validation loss.

- **Visualization of the model's loss and accuracy over the training process:**



- **Model Accuracies:**
  1) Training: around 70%
  2) Validation: 68.6%
  3) Test data: 60.1 %

Task 2.3 :

- **Key Characteristics of the Pigmented Skin Lesions Dataset vs. the Models:**
  The pigmented skin lesions dataset poses specific challenges such as high variability in lesion appearance due to differences in shape, size, color, and texture. VGG16, pretrained on a diverse set of ImageNet pictures, offers a broad understanding of visual features but might lack specialization in medical imagery. Its deep architecture can capture complex patterns, which could be advantageous for distinguishing subtle differences in skin lesions. Conversely, a custom model, while not benefiting from pre-existing knowledge, can be tailored to emphasize features unique to skin lesions, potentially leading to better performance with less computational overhead. The custom model may also be more

adaptable to specific augmentations or preprocessing steps that are relevant to medical images, such as normalizing for variations in skin tone or lighting conditions.

- **Complexity of the Data:**
  The complexity and diversity of pigmented skin lesion images can make using a pretrained model like VGG16 appealing because it has been trained to recognize a wide range of features that could be pertinent to identifying lesions. However, the specificity required for medical diagnosis means that a custom model could potentially be designed to focus on medically relevant features that might be underrepresented in generic datasets like ImageNet. If the variability within classes is significant, a custom model could be trained to pay more attention to the most discriminative features of the lesions, which might be overlooked by a general-purpose feature extractor like VGG16.

- **Selecting a Model Considering the Constraints:**
  Considering the trade-offs, if the dataset of pigmented skin lesions is limited or highly specific, a custom neural network might be the preferred choice. Custom models can be designed to be less data-hungry and more focused on the task-specific features, potentially leading to better performance with the available data. On the other hand, if a large dataset is available, foundation models like VGG16 could be leveraged for their powerful feature extraction capabilities. For a domain as specialized as pigmented skin lesions, the choice would ideally also depend on the domain knowledge integrated into the model design, which could range from specialized data augmentations to custom layer architectures.

- **Feature-Based Model for Pigmented Skin Lesions vs. Wafer Maps:**
  Extracting features from pigmented skin lesion images is typically more challenging than from wafer maps due to the complex and varied nature of biological data. Skin lesions can have irregular patterns, varied textures, and colors, and they exist in different shapes and sizes, often on varied skin tones. In contrast, wafer maps tend to have more regular patterns and are usually analyzed in a controlled environment, reducing variability. For skin lesions, advanced techniques such as deep learning might be necessary to capture the intricate details required for accurate classification, whereas wafer maps might be effectively analyzed using more traditional image processing techniques due to their structured nature.

Task 3.1 :

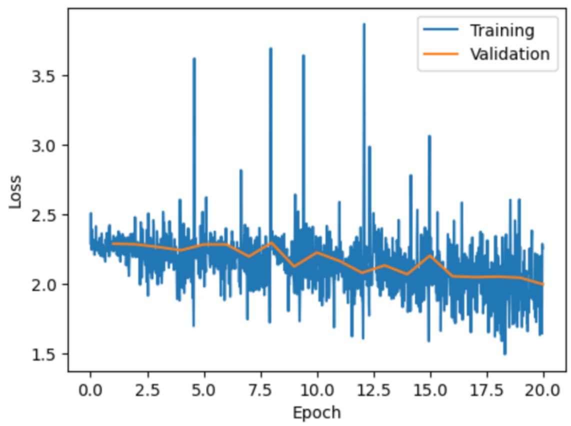
- **Data Preprocessing Steps and Data Split Ratio:**
  For the CalTech-256 dataset, the preprocessing included padding images to create a square shape, converting them to tensors, resizing them to 224x224 pixels, and applying a

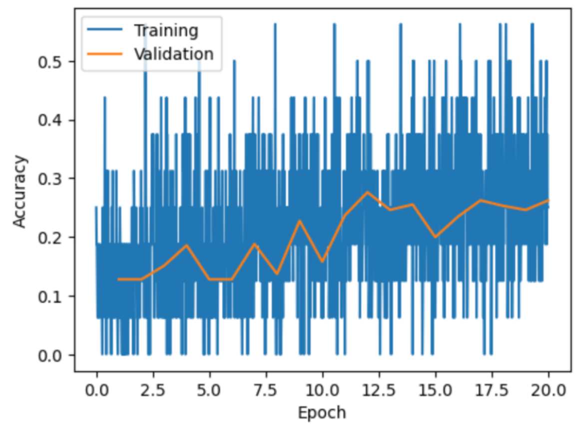
standard transformation pipeline without normalization. This standardization is crucial for preparing the data for a neural network that expects a specific input size and format. The data was split into training and validation sets using a 70-30 ratio, providing a substantial dataset for training while retaining a significant amount for validation. This split helps to train the model effectively while also allowing for a robust evaluation of its performance.

- **Structure of the Model:**
  The model, named CalTechNetwork, is a custom neural network that comprises several convolutional layers with increasing numbers of filters (32, 64, 128, 128), each followed by a max-pooling layer. The network then transitions to fully connected layers, with the final layer having an output size that matches the number of classes in the dataset. Dropout is applied after each convolutional layer and the first fully connected layer to prevent overfitting. The custom architecture is designed to capture the hierarchical features of the CalTech-256 images, which can range from simple edges to complex shapes, and is tuned specifically for the task of image classification within this dataset.

- **Optimization Parameters:**
  The custom model was optimized using Stochastic Gradient Descent (SGD) with a learning rate of 0.1, which is relatively high, suggesting confidence in the model's ability to learn quickly without overshooting the minima. No momentum or weight decay parameters are mentioned, indicating a straightforward application of SGD. This approach can be effective for datasets with a large variety of classes, as it helps to explore the solution space rapidly. However, care must be taken to adjust the learning rate as needed to prevent the model from becoming stuck in local minima or failing to converge. The model is trained for 20 epochs, providing a balance between sufficient training time and computational efficiency.

- **Visualization of the model's loss and accuracy over the training process:**





- **Model Accuracies:**
    1) Training: around 45%
    2) Validation: 26.2%
    3) Test data: 94 %

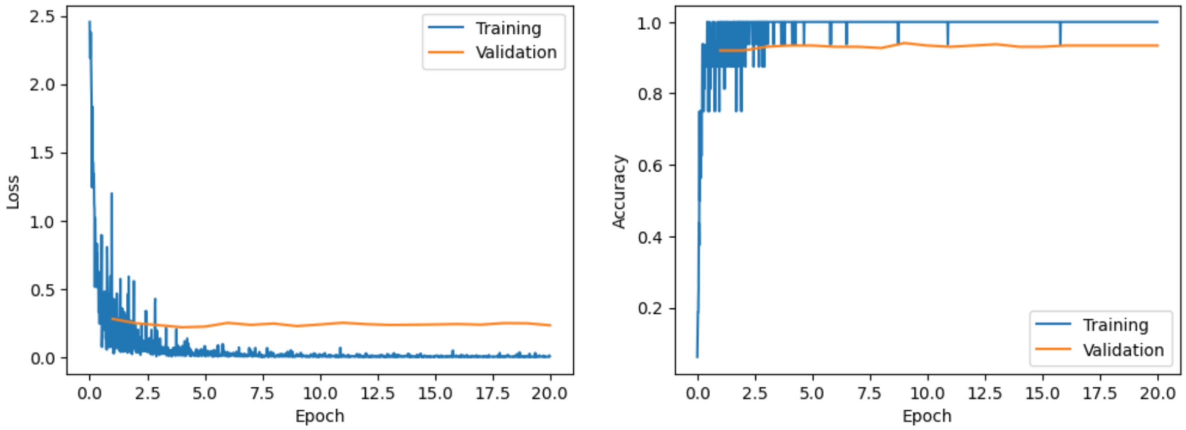
Task 3.2 :

- **Data Preprocessing Steps and Data Split Ratio:**
  The preprocessing for the CalTech-256 dataset involved transforming the images to ensure they were square by padding, converting them to tensors, resizing to 224x224 pixels (the input size required by VGG16), and normalizing them with the mean and standard deviation values used with ImageNet. The split ratio between the training and validation sets was 80-20, allocating a majority of the images for training to maximize the learning potential of the model, while still reserving a substantial portion for validation to assess the model's generalization to new data.

- **Optimization Parameters:**
  The VGG16 model, adapted for the CalTech-256 dataset, was fine-tuned using Stochastic Gradient Descent (SGD) with a learning rate of 0.001 and a momentum of 0.9. These optimization parameters were selected to gradually adjust the weights of the network to the new dataset, with a low learning rate to avoid overwriting the pre-learned features from ImageNet too quickly, and a high momentum to help accelerate the convergence and navigate the complex optimization landscape efficiently. The fine-tuning process involved the entire model, but with the convolutional base layers frozen to retain their

learned features and only the fully connected layers being trained to adapt to the new dataset.

- **Visualization of the model's loss and accuracy over the training process:**



- **Model Accuracies:**
  1) Training: around 90%
  2) Validation: 93.4%
  3) Test data: 94 %

Task 3.3 :

- **Key Characteristics of the CalTech-256 Dataset vs. the Models:**
  The CalTech-256 dataset, comprised of natural images of objects, presents a significant challenge due to the high intra-class variability and inter-class similarity. Objects within the same category can appear very different, while objects from different categories can appear similar. A pretrained model like VGG16 comes with an extensive knowledge base learned from ImageNet, which covers a wide range of natural images, potentially providing a solid starting point for feature extraction. It may have an advantage in identifying the diverse set of features present in the CalTech-256 images. However, a custom model can be designed to focus on the specific attributes of the CalTech-256 dataset, possibly requiring fewer resources and providing a more tailored solution, albeit without the benefit of transfer learning.

- **Complexity of the Data:**
  The diversity of image sizes and the high variability within classes in the CalTech-256 dataset may make VGG16 a suitable choice due to its depth and capacity to learn complex features. However, this can also lead to overfitting if not managed correctly. A custom model, on the other hand, allows for the flexibility to design an architecture that

matches the complexity of the dataset, potentially optimizing the number of parameters to reduce overfitting while still capturing the essential features necessary for classification.

- **Selecting a Model Considering the Constraints:**
  When considering data availability and characteristics, along with the applicability of the models to the domain of natural images of common objects, foundation models like VGG16 are often the preferred choice due to their proven capability in image classification tasks. They provide a robust starting point for feature extraction, leveraging the extensive training on ImageNet. However, if computational resources are limited or if there is a need for a model with fewer parameters for faster inference, a custom neural network may be preferred. The custom network can be fine-tuned more precisely to the CalTech-256 dataset, provided that there is sufficient domain knowledge to guide the architecture and training process.

- **Feature Extraction for Different Data Types:**
  Extracting features from wafer maps is generally a more straightforward task due to their structured and consistent patterns. Pigmented skin lesions and natural images of objects are significantly more challenging. Skin lesions require careful analysis of texture, color, and boundary irregularities, which are subtle and highly variable. Natural images encompass an even broader range of features, requiring a model to understand context, object relations, and a vast array of shapes and textures. A feature-based model for skin lesions or natural images would need to be much more sophisticated and robust, potentially leveraging deep learning techniques to capture the hierarchical and abstract features that are crucial for accurate classification.

Task 4:

- **Main Differences or Uniqueness for Each of the Three Datasets:**
  The WaferMap dataset typically consists of structured patterns that are often geometric and regular, making it distinct due to the predictable nature of defects and the controlled manufacturing environment. This results in a dataset with lower intra-class variability and well-defined features.
  The HAM10000 (Pigmented Skin Lesions) dataset is characterized by high variability in lesion appearance, including irregular shapes, varied colors, and textures, which may also be influenced by factors like skin tone and image capture conditions. These images require careful analysis of subtle medical features that are not commonly found in standard image datasets.
  The CalTech-256 dataset, containing natural images of objects, presents high intra-class variability and inter-class similarities. The images are more complex, with varied

backgrounds and scales, and include a wide range of common objects that can appear in many different forms, positions, and environmental conditions.

- **Circumstances for Choosing a Training Method:**
  Feature-based models would be chosen for the WaferMap dataset or other similarly structured datasets where features are more regular and can be engineered without deep learning. Such models require less computational power and can be more interpretable, which is often important in manufacturing settings.
  Custom neural networks might be preferred in situations with limited data, when domain-specific features need to be captured, or when computational resources are a constraint. For instance, if the goal is to fine-tune the model for specific attributes found in the skin lesion images, which may not be well-represented in pre-trained models.
  Foundation models like VGG16 would be ideal for the CalTech-256 dataset, leveraging the power of transfer learning where a vast and diverse set of features from ImageNet can be fine-tuned to the task at hand. Such models are suitable when a large, varied dataset is available, and there is a need to capture complex patterns.

- **Easier Dataset to Classify:**
  The term "easier" in the context of classification could refer to a dataset that requires less complex models to achieve high accuracy, has lower intra-class variability and inter-class similarity, and presents patterns that are more distinct and separable by standard feature extraction methods.
  Given these criteria, the WaferMap dataset might be considered "easier" to classify than the other two. The structured nature of the defects in wafer maps, which tend to have consistent patterns, can often be classified using simpler models or traditional machine learning algorithms with handcrafted features. In contrast, the HAM10000 and CalTech-256 datasets, with their high variability and less structured data, likely require more complex models such as deep learning to achieve high accuracy, which can be more challenging to train and fine-tune.