MileStone 1 - Initial submission
Deep Learning and Software Engineering

Flower Finder

Team - Group 10 :
ALAZARD Camille
BRITZELMAIER Felix
WALLISER Fabian

1) Context

• **Application Overview**

The Flower Finder is designed to help users identify flowers through two options: uploading an image or taking a new one using their device's camera. Users can quickly find out what type of flower they are looking at, making the application useful for those unfamiliar with flower names. The images are classified using a deep learning model. Once the flower is identified, the system automatically sorts the flower into its corresponding directory for easier access and categorization.

• **Hiking User Example**

The Flower Finder is perfect for anyone interested in flowers, especially hikers. Imagine a hiker encountering an unfamiliar flower during their hike. They can simply take a picture or upload an image through the app. The system will identify the flower species and automatically place it in the correct directory, allowing the hiker to access it for future reference and research.

2) High-Level Description of the Model

After looking at multiple different models, we chose to use the Vision Transformer (ViT) for the task at hand. The main reasons for this decision were its very good performance in image classification and the high amount of data it was trained on. To explain this further, we will show some other possible models and explain the reasons we did not choose them.

• **Other considered models**

ResNet-152 is a deep convolutional neural network widely used in computer vision classification. It is highly accurate due to its depth, but the significant amount of computational power required was a major drawback. Its depth could also lead to overfitting on small datasets like ours.

Swin Transformers is a vision transformer-based model that divides the image into smaller patches and applies self-attention mechanisms, making it a general-purpose backbone for image classification. However, it requires a high amount of computational power and complexity, which is excessive for our relatively simple task.

Convolutional Neural Network (CNN) is a classical model widely used for image classification due to its simplicity and customizability. However, its limited ability to capture long-range dependencies and global features makes it less suitable for more complex datasets, even for simpler tasks like flower classification.

EfficientNetB3 is a well-rounded model that scales the number of layers and parameters to optimize accuracy and efficiency. However, it requires extensive parameter tuning and computational resources, which is not ideal for our project scope.

• **Justification for Choosing Vision Transformer (ViT)**

As shown by the examples above, many models require either extensive datasets, high computational power, or much tuning. The Vision Transformer, while slightly resource-intensive, is more suitable for our task due to its strong performance on various image classifications. There also are many examples of the model being fine-tuned and therefore the project success should be granted.

Model Card - Flower classification

1) Model Details

• **Person or organization developing the model:** Developed by Google Research and first introduced by Dosovitskiy et al. The weights were converted from the timm repository by Ross Wightman.
• **Model date:** Released in 2020.
• **Model version:** Patch 16
• **Model type:** Vision Transformer (ViT), a Transformer encoder model designed for image recognition tasks.
• **Training algorithms and parameters:** The ViT model is trained by using supervised learning on ImageNet-21k. It uses the Transformer architecture which was originally used for NLP tasks and processes images as a sequence of patches.
• **Paper:** "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" by Dosovitskiy et al.
• **Citation:** See BibTeX citation at the end of the card.
• **License:** Apache 2.0 License.
• **Where to send questions or comments about the model:** Contact the Hugging Face community, where the model is hosted.

2) Intended Use

• **Primary intended uses**
This model is intended for image classification tasks. It has been pre-trained on a large dataset (ImageNet-21k), which makes it a flexible base model for various downstream tasks.

• **Primary intended users**
Researchers, Developers, and any type of machine learning practitioners who want to use a Transformer-based model for image recognition tasks.

• **Out-of-scope use cases**
This model is not fine-tuned for tasks outside image classification. The pre-trained weights are not designed for image generation or object detection, and applying the model to other domains without appropriate fine-tuning may lead to suboptimal results.

3) Factors

• **Relevant factors**
Hardware factors of camera and lens type as well as environmental factors of lighting and humidity to prevent blurred or distorted images can affect the effectiveness of the model in extracting important features.

• **Evaluation factors**
The performance varies with the dataset size, diversity, and the distribution of the existing training images across the categories.

4) Metrics

• **Model performance measures**
Standard image classification metrics such as accuracy, top-1, and top-5 accuracy were used for evaluation in the paper.

• **Decision thresholds**
Not applicable as this model gives scores for each class, which can be turned into probabilities using a softmax function. These probabilities show how likely it is that an image belongs to each class.

• **Variation approaches**
Larger model variants (such as ViT-Large) and higher input resolutions (384x384) can improve performance. Fine-tuning also leads to better results for domain-specific tasks.

5) Evaluation Data

• **Datasets**
The ViT model was pre-trained on ImageNet-21k, a large-scale dataset with 14 million images and 21,843 categories (test data split).

• **Motivation**
ImageNet-21k is one of the largest and most diverse image datasets available, making it suitable for pre-training models that can generalize well to various downstream tasks.

• **Preprocessing**
Images are resized/rescaled to the same resolution (224x224) and normalized across the RGB channels with mean (0.5, 0.5, 0.5) and standard deviation (0.5, 0.5, 0.5). The exact details about the preprocessing can be found here.

6) Training Data

Training data was sourced from the same dataset as the evaluation data and was preprocessed in the same way (training data split).

7) Quantitative Analyses

• **Unitary results:** The ViT model achieved strong performance on several image classification benchmarks. For specific accuracy values, refer to tables 2 and 5 in the original paper.

8) Ethical Considerations

• **Bias and Fairness**
Since the model is trained on ImageNet, which contains images from various online sources, there may be inherent biases in the data, such as underrepresentation of certain demographics or overrepresentation of Western cultural elements. These biases may affect the model's performance in real-world applications.

• **Potential Misuse**
This model is intended for image classification tasks and should not be used in critical applications such as medical diagnosis, surveillance, or applications that can negatively impact individuals' privacy or security without proper oversight and testing.

9) Caveats and Recommendations

• This model performs best when fine-tuned on a task-specific dataset.
• Larger input resolutions (e.g., 384x384) can improve classification accuracy but come at the cost of increased computational resources.

10) BibTeX Entry and Citation Information

```
@misc{wu2020visual,
      title={Visual Transformers: Token-based Image Representation and Processing for Computer Vision},
      author={Bichen Wu and Chenfeng Xu and Xiaoliang Dai and Alvin Wan and Peizhao Zhang and Zhicheng Yan and Masayoshi Tomizuka and Joseph Gonzalez and Kurt Keutzer and Peter Vajda},
      year={2020},
      eprint={2006.03677},
      archivePrefix={arXiv},
      primaryClass={cs.CV}
}
@inproceedings{deng2009imagenet,
  title={Imagenet: A large-scale hierarchical image database},
  author={Deng, Jia and Dong, Wei and Socher, Richard and Li, Li-Jia and Li, Kai and Fei-Fei, Li},
  booktitle={2009 IEEE conference on computer vision and pattern recognition},
  pages={248--255},
  year={2009},
  organization={Ieee}
}
```

Data Card - 102 Flower classification

1) Dataset Description

The **102 Category Flower Dataset** consists of **8,189 images** of flowers belonging to **102 different species**, chosen for their common occurrence in the United Kingdom. Each flower category is composed of 40 to 258 flower pictures. The dataset was created to assist in the development and evaluation of flower **classification** systems. Some categories have flowers highly similar to others, introducing additional challenges for classification and allowing to evaluate the performances of the classification device better.

The dataset also includes segmentations of the images and chi-squared (χ^2) distances used for feature matching.

• **Homepage:** https://www.robots.ox.ac.uk/~vgg/data/flowers/102/index.html
• **Repository :** https://www.robots.ox.ac.uk/~vgg/data/flowers/102/102flowers.tgz (329 Mo)
• **Languages :** English (flowers labels)
• **Version :** 1.1

• **Paper:** Automated flower classification over a large number of classes (2008)
• **Point of Contact:** - Maria-Elena Nilsback (men@robots.ox.ac.uk), University of Oxford
- Andrew Zisserman (az@robots.ox.ac.uk), University of Oxford

2) Dataset Structure

• **Data Fields**
The dataset contains the following fields :

1) **Images :** Images of the flowers, in .jpg format, belonging to one of the 102 categories. They are numbered from 1 to 8,189 and this number serves as an unique id. They are the primary input for classification tasks and vary in terms of size, lighting, and pose

2) **Segmentations**: These are the segmentation masks of the flower images, which can be used for segmentation-based tasks. *(Not used for our project)*

3) **Chi-squared (χ^2) distances**: These are used to represent feature distances between images for comparative and classification purposes. *(Not used for our project)*

4) **Labels**: The imagelabels.mat file associate each picture with its numerical label between 1 and 102, representing each flower category.

• **Data Split**
The dataset is already splitted according to the file setid.mat (see Annexe - Class Distribution for more information) in **12.5 % training set**, **12.5% validation set** and **75% test set**.

3) Dataset Creation

• **Curation Rationale**

The intended tasks supported by the dataset include flower classification and recognition, as well as testing algorithms for feature extraction and pattern recognition.

The challenges that interested Maria-Elena Nilsback and Andrew Zisserman more particularly when creating the dataset were :

1) The large **similarity between classes**.
2) The large **variation within classes** due to flowers being non-rigid objects that can deform in many ways.
3) The **large number of classes** with 103 classes where previous papers used from 10 to 30 classes.

• **Source Data**

According to the paper, most of the images were collected from the web and small number of images were acquired by Maria-Elena Nilsback and Andrew Zisserman taking the pictures themselves.

• **Annotations**

The relation between numerical labels and the names of the flower classes were missing. As such, the project team added a new annotation, named labels.py, by hand. This file associates the numerical label to the flower species names as string labels.
Example : { 1: 'pink primrose', 2 : 'hard-leaved pocket orchid', 3 : 'canterbury bells', … }

Methodology :

1) Searching for the id of the picture illustrating the class in the class list on the document of reference
2) Searching for the numerical label in the imagelabels.mat file from the picture id
3) Writing the association between the numerical label and the name of the flower specie in labels.py

The documents used as a reference to do the annotations is the list of classes illustrated with a picture from the dataset : https://www.robots.ox.ac.uk/~vgg/data/flowers/102/categories.html.

• **Personal and Sensitive Information**

The dataset only contains pictures of flowers without metadata such as authors, location, etc… so there is no sensitive information.

4) Considerations for Using the Data

• **Discussion of Biases**

The dataset is limited **geographically** : it only contains flowers commonly found in the United Kingdom and doesn't represent flowers found in other parts of the world.
Plus, the image variations (pose, lighting and scale) could create biases when using the dataset for developing classification algorithms that may not generalize well to **real-world scenarios**.

• **Other Known Limitations**

One limitation of this dataset is the **small number of images** per category in the **training** and **validation** dataset. Because there can be 40 to 258 pictures of flowers by classes, only 10 pictures by classes are used in the training and validation datasets to avoid a category to be overrepresented in one of these two sets. It could badly affect the performance of machine learning models trained on it.

5) Additional Information

• **Dataset Curators**

The dataset has been constituted by Maria-Elena Nilsback and Andrew Zisserman from the University of Oxford with the assistance of Radhika Desikan, Liz Hodgson and Kath Steward, the experts who assisted in labelling the flower classes.

There work was funded by the EC Marie-Curie Training Network VISIONTRAIN, Microsoft Research and the Royal Academy of Engineering.

• **License :** Unknown

But the paper is associated with the following note regarding copyright :

>

• **Contributions** : Thanks to Maria-Elena Nilsback and Andrew Zisserman from the University of Oxford for creating this dataset.

• **Citation Information :**

```
@InProceedings{Nilsback08,
  author      = "Maria-Elena Nilsback and Andrew Zisserman",
  title       = "Automated Flower Classification over a Large Number of Classes",
  booktitle   = "Indian Conference on Computer Vision, Graphics and Image Processing",
  month       = "Dec",
  year        = "2008",
}
```

Annexe - Class Distribution

	Train	Validation	Test	Total
Pink Primrose	10	10	20	**40**
Hard-leaved pocket orchid	10	10	40	**60**
Canterbury Bells	10	10	20	**40**
Sweet Pea	10	10	36	**56**
English Marigold	10	10	45	**65**
Tiger Lily	10	10	25	**45**
Moon Orchid	10	10	20	**40**
Bird of Paradise	10	10	65	**85**
Monkshood	10	10	26	**46**
Globe Thistle	10	10	25	**45**
Snapdragon	10	10	67	**87**
Colt's Foot	10	10	67	**87**
King Protea	10	10	29	**49**
Spear Thistle	10	10	28	**48**
Yellow Iris	10	10	29	**49**
Globe-flower	10	10	21	**41**
Purple Coneflower	10	10	65	**85**
Peruvian Lily	10	10	62	**82**
Balloon Flower	10	10	29	**49**
Giant White Arum Lily	10	10	36	**56**
Fire Lily	10	10	20	**40**
Pincushion Flower	10	10	39	**59**
Fritillary	10	10	71	**91**
Red Ginger	10	10	22	**42**
Grape Hyacinth	10	10	21	**41**
Corn Poppy	10	10	21	**41**
Prince of Wales Feathers	10	10	20	**40**
Stemless Gentian	10	10	46	**66**
Artichoke	10	10	58	**78**
Sweet William	10	10	65	**85**

Carnation	10	10	32	**52**
Garden Phlox	10	10	25	**45**
Love in the Mist	10	10	26	**46**
Mexican Aster	10	10	20	**40**
Alpine Sea Holly	10	10	23	**43**
Ruby-lipped Cattleya	10	10	55	**75**
Cape Flower	10	10	88	**108**
Great Masterwort	10	10	36	**56**
Siam Tulip	10	10	21	**41**
Lenten Rose	10	10	47	**67**
Barbeton Daisy	10	10	107	**127**
Daffodil	10	10	39	**59**
Sword Lily	10	10	110	**130**
Poinsettia	10	10	73	**93**
Bolero Deep Blue	10	10	20	**40**
Wallflower	10	10	176	**196**
Marigold	10	10	47	**67**
Buttercup	10	10	51	**71**
Oxeye Daisy	10	10	29	**49**
Common Dandelion	10	10	72	**92**
Petunia	10	10	238	**258**
Wild Pansy	10	10	65	**85**
Primula	10	10	73	**93**
Sunflower	10	10	41	**61**
Pelargonium	10	10	51	**71**
Bishop of llandaff	10	10	89	**109**
Gaura	10	10	47	**67**
Geranium	10	10	94	**114**
Orange Dahlia	10	10	47	**67**
Pink-yellow Dahlia	10	10	89	**109**
Cautleya Spicata	10	10	30	**50**
Japanese Anemone	10	10	35	**55**
Black-eyed Susan	10	10	34	**54**

Silverbush	10	10	32	**52**
Californian Poppy	10	10	82	**102**
Osteospermum	10	10	41	**61**
Spring Crocus	10	10	22	**42**
Bearded Iris	10	10	34	**54**
Windflower	10	10	34	**54**
Tree Poppy	10	10	42	**62**
Gazania	10	10	58	**78**
Azalea	10	10	76	**96**
Water Lily	10	10	174	**194**
Rose	10	10	151	**171**
Thorn Apple	10	10	100	**120**
Morning Glory	10	10	87	**107**
Passion Flower	10	10	231	**251**
Lotus	10	10	117	**137**
Toad Lily	10	10	21	**41**
Anthurium	10	10	85	**105**
Frangipani	10	10	146	**166**
Clematis	10	10	92	**112**
Hibiscus	10	10	111	**131**
Columbine	10	10	66	**86**
Desert-rose	10	10	43	**63**
Tree Mallow	10	10	38	**58**
Magnolia	10	10	43	**63**
Cyclamen	10	10	134	**154**
Watercress	10	10	164	**184**
Canna Lily	10	10	62	**82**
Hippeastrum	10	10	56	**76**
Bee Balm	10	10	46	**66**
Ball Moss	10	10	26	**46**
Foxglove	10	10	142	**162**
Bougainvillea	10	10	108	**128**
Camellia	10	10	71	**91**

Mallow	10	10	46	**66**
Mexican Petunia	10	10	62	**82**
Bromelia	10	10	43	**63**
Blanket Flower	10	10	29	**49**
Trumpet Creeper	10	10	38	**58**
Blackberry Lily	10	10	28	**48**
All input pictures	**1,020**	**1,020**	**6,149**	**8,189**