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**