

Курсовой проект по дисциплине МО “Классификация цветковых растений по изображению”

Работу выполнил студент Жилкин Федор Игоревич

Санкт-Петербург, 2022 г.

Постановка задачи

В базе данных [The Plant List](#) содержится 304 419 признанных видовых названий цветковых растений, принадлежащих к 405 семействам и 14 559 родам, при этом общее **количество видов** цветковых растений оценивается в 352 000. Некоторые из этих цветковых растений могут быть сорными, некоторые вызывают аллергию, а некоторые даже ядовиты для человека.

Различные цветковые растения могут быть использованы для украшения ландшафта, при дизайне одежды, во время изготовления косметических средств.

Становится очевидно, что **необходимо уметь различать разные цветковые растения друг от друга**. Один из способов это сделать -- использовать алгоритмы компьютерного зрения для определения растения по его фотографии.

Исходя из всего вышеперечисленного, составим цель и задачи исследования.

Цель исследования:

Построить алгоритм, способный предсказать цветковое растение по его фотографии.

Задачи:

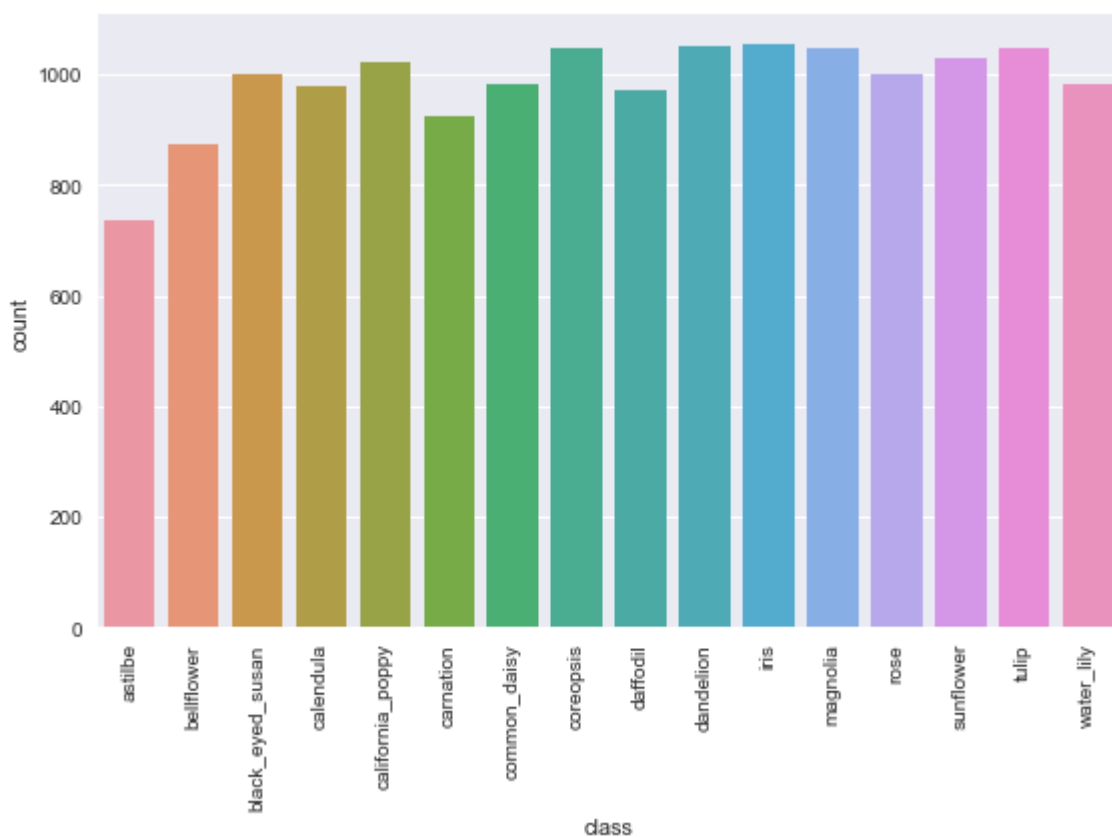
- Найти подходящий набор данных (далее датасет) и описать его характеристики
- Описать использованные методы компьютерного зрения, позволяющие решить задачу классификации растений
- Обучить алгоритмы на выбранном датасете
- Провалидировать каждый из алгоритмов на тестовом множестве изображений цветковых растений

Набор данных 🌸 | Flowers

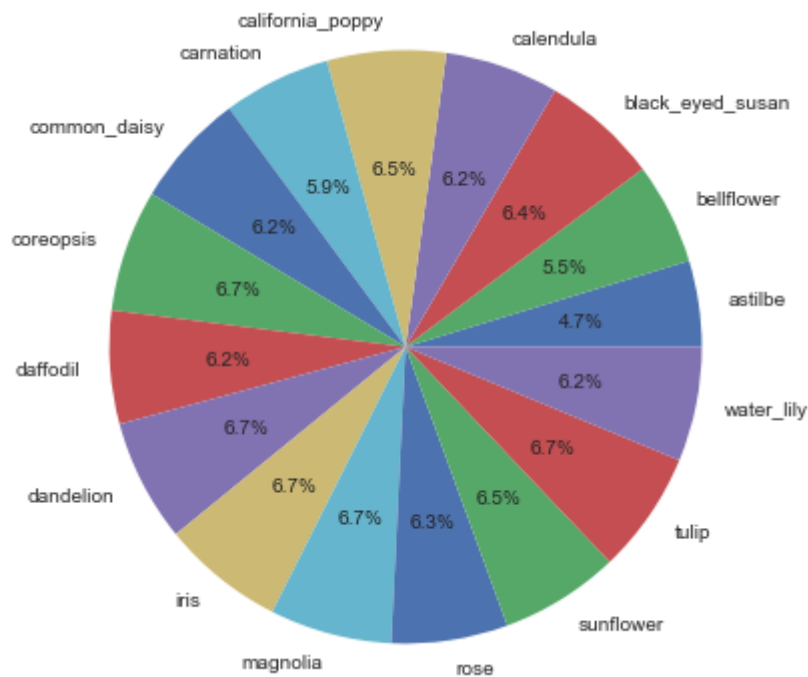
Набор данных 🌸 [Flowers](#) содержит 15700 изображений растений, разделенных на 16 классов:

- Астильба
- Колокольчик
- Маргаритка глориоза
- Календула
- Эшшольция калифорнийская
- Гвоздика
- Маргаритка многолетняя
- Кореопсис
- Нарцисс
- Одуванчик
- Ирис
- Магнолия
- Роза
- Подсолнух
- Тюльпан
- Водная Лилия.

Гистограмма количества изображений в каждом классе:

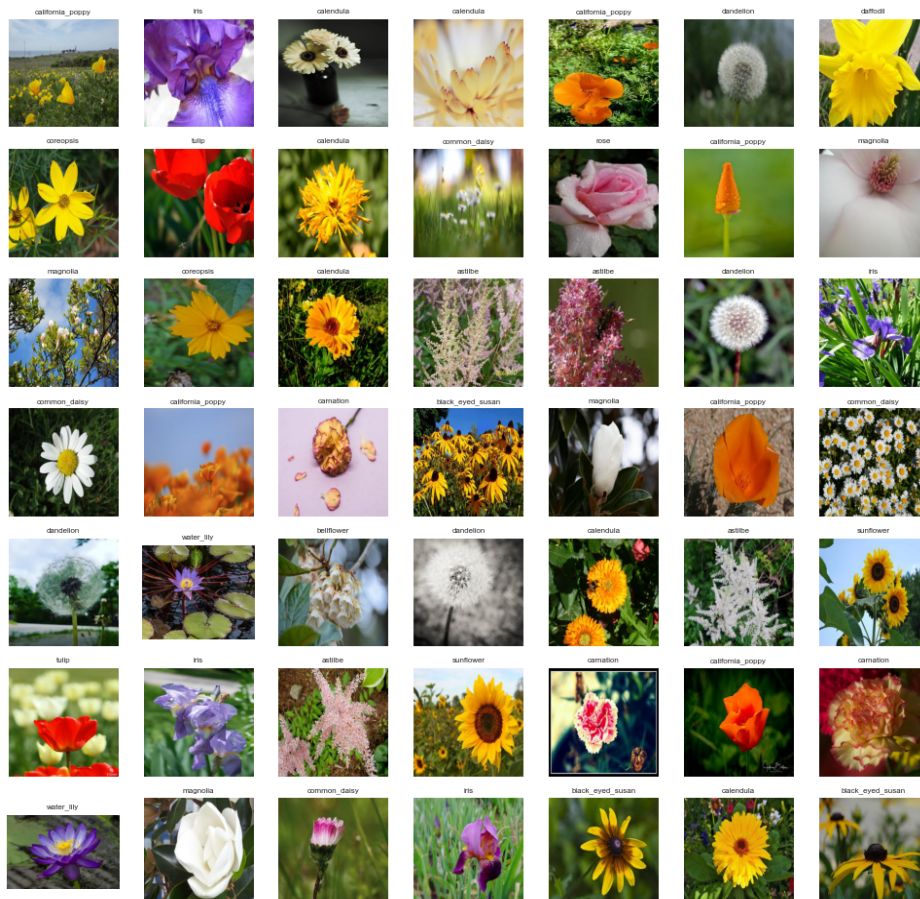


Доля изображений в каждом классе:



Используя графические средства представления данных, нетрудно заметить, что разброс в количестве изображений невелик, поэтому данные будем считать **сбалансированными**.

Пример изображений в датасете:



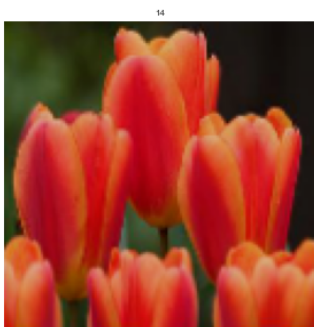
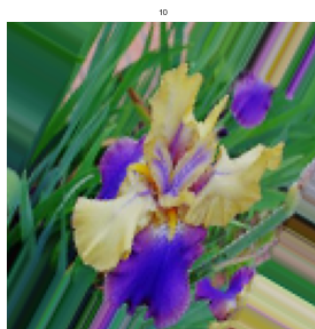
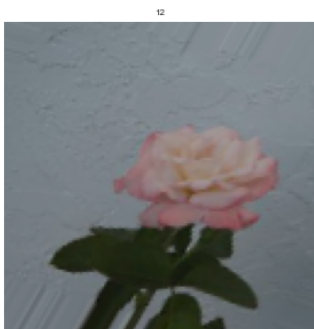
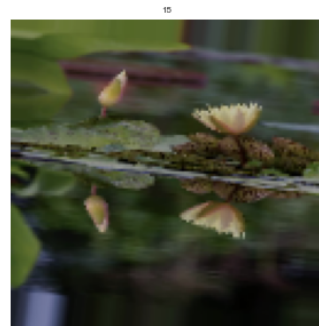
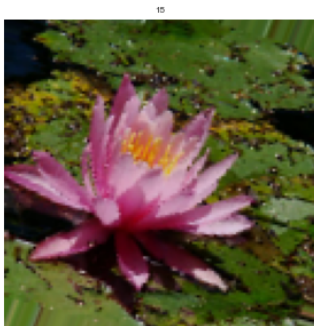
Подготовка данных

Под [аугментацией](#) данных понимается увеличение выборки данных для обучения через модификацию существующих данных.

Проведем аугментацию исходных данных с использованием следующих преобразований:

- rotation_range
- width_shift_range
- height_shift_range
- shear_range
- zoom_range
- horizontal_flip
- vertical_flip
- brightness_range

Пример изображений после аугментации:



Классификация изображений

Для решения задачи выберем модели, зарекомендовавшие себя в классификации изображений на разных бенчмарках и датасетах:

- DenseNet121
- EfficientNet-B0
- VGG16

Densely Connected Convolutional Network

[DenseNet121](#) была предложена в 2017 году. Успех ResNet (Deep Residual Network) позволил предположить, что укороченное соединение в CNN позволяет обучать более глубокие и точные модели. Авторы проанализировали это наблюдение и представили компактно соединенный (dense) блок, который соединяет каждый слой с каждым другим слоем. Важно отметить, что, в отличие от ResNet, признаки («фичи») прежде чем они будут переданы в следующий слой не суммируются, а конкатенируются (объединяются, channel-wise concatenation) в единый тензор. При этом количество параметров сети DenseNet намного меньше, чем у сетей с такой же точностью работы. Авторы [утверждают](#), что DenseNet работает особенно хорошо на малых наборах данных.

EfficientNet

[EfficientNet-B0](#) — класс моделей, который получился из изучения масштабирования (scaling) моделей и балансирования между собой глубины и ширины (количества каналов) сети, а также разрешения изображений в сети. Авторы [статьи](#) предлагают новый метод составного масштабирования (compound scaling method), который равномерно масштабирует глубину/ширину/разрешение с фиксированными пропорциями между ними. Из существующего метода под названием «Neural Architecture Search» для автоматического создания новых сетей и своего собственного метода масштабирования авторы получают новый класс моделей под названием EfficientNets.

VGG16

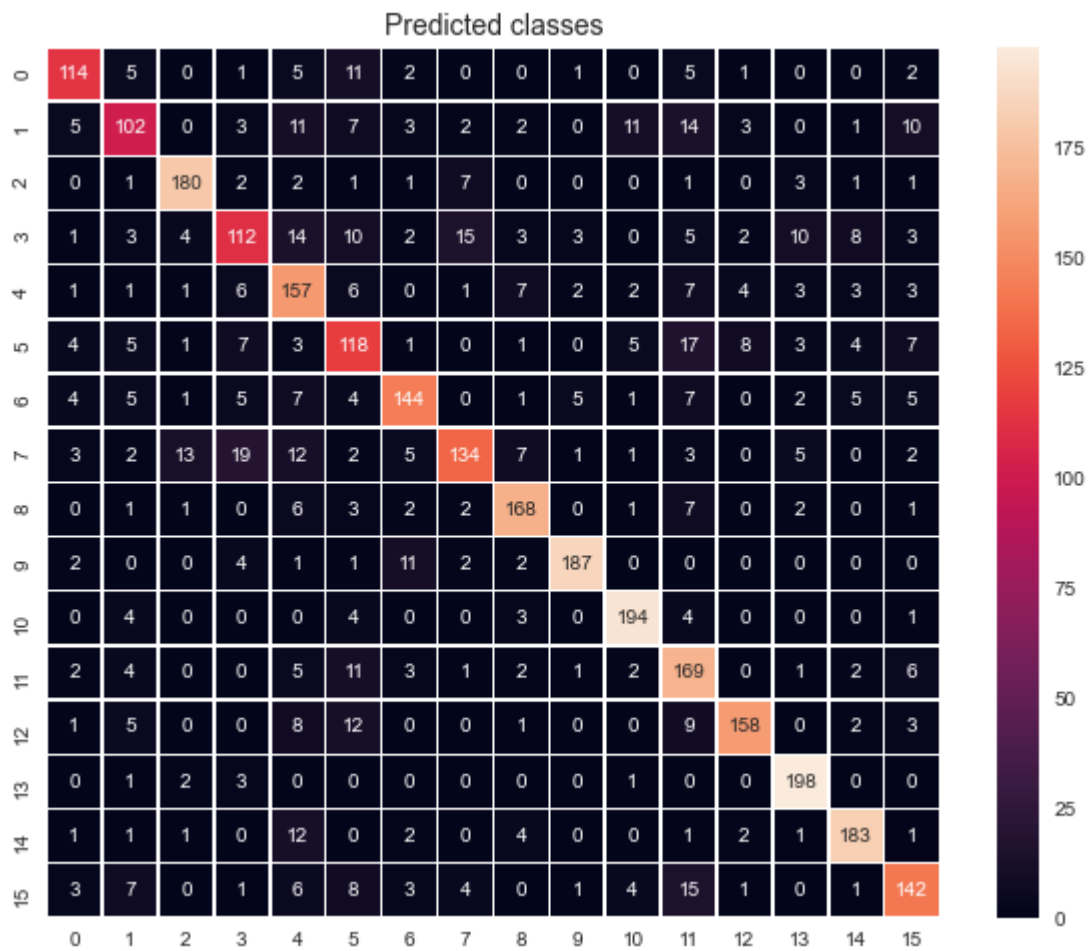
[VGG16](#) — модель сверточной нейронной сети, предложенная К. Simonyan и А. Zisserman из Оксфордского университета в статье “Very Deep Convolutional Networks for Large-Scale Image Recognition”. Модель достигает точности 92.7% — топ-5, при тестировании на ImageNet в задаче распознавания объектов на изображении. Этот датасет состоит из более чем 14 миллионов изображений, принадлежащих к 1000 классам.

Первая часть решения будет заключаться в использовании метода **Transfer Learning**. В качестве весов возьмем параметры imagenet. Затем мы обучаем модель на наших данных. Это позволит классификатору лучше выделять отличительные особенности (features).

Вторая часть решения заключается в использовании метода **Fine Tuning**. Для этого удалим 1 последний слой и обучим модель еще раз.

При оценивании качества работы моделей будем использовать метрику f1-score. Дополнительно построим матрицу сопряженности (Confusion Matrix) и графики (Training-Validation Acc, Training-Validation loss).





Оценка результатов и выводы

Исходя из проведенных тестов, получаем следующие значения метрик:

	Precision	Recall	f1-score
DenseNet121	0.87	0.87	0.87
EfficientNet-B0	0.84	0.86	0.85
VGG16	0.78	0.77	0.77

Вывод

Наивысшие значения метрик достигаются с использованием архитектуры **DenseNet121**. Также неплохо себя показала модель **EfficientNet-B0**. Для следующих исследований можно попробовать обучить **EfficientNet** модели, но с бОльшим количеством параметров (**EfficientNet-B2, -B3, -B7**).

Код исследования доступен в GitHub: <https://github.com/Feodoros/FlowerRecognition>