**Практическая работа 1.** **Алгоритмы кластеризации в задачах обработки изображений**

**Цель**:

Изучить применение моделей кластеризации, а также библиотек обработки изображений

**Описание задания**:

В домашнем задании нужно сократить число цветов в палитре[изображения](https://www.firestock.ru/wp-content/uploads/2018/03/YArkaya-klumba-Bright-flower-bed-6000x4000-700x466.jpg) [url: <https://www.firestock.ru/wp-content/uploads/2018/03/YArkaya-klumba-Bright-flower-bed-6000x4000-700x466.jpg> ]. Картинку для выполнения работы можно выбрать любую, главное условие – наличие на ней разных цветов, для того, чтобы результат работы моделей был заметен.

Для выполнения работы необходимо выделить кластеры в пространстве RGB, объекты соответствуют пикселям изображения. После выделения кластеров все пиксели, отнесенные в один кластер, заполняются одним цветом. Цвет – центроид соответствующего кластера.

**Этапы работы**:

1. Получите данные и загрузите их в рабочую среду (Jupyter Notebook или другую).
2. Для отображения картинки воспользуйтесь вспомогательной функцией:

import matplotlib.pyplot as plt

import cv2

import skimage

def draw\_picture(image, title, bgr=False):

b, g, r = cv2.split(image) # по умолчанию cv2 почему-то отдает цвета в порядке BGR вместо RGB

new\_image = cv2.merge([r, g, b])

plt.figure(figsize=(7, 5))

plt.axis(‘off’)

plt.imshow(new\_image)

plt.title(title)

plt.show()

1. Для загрузки картинки можно использовать код:

img = cv.imread(path)

1. Реализуйте три модели кластеризации:

a) [KMeans](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html" \t "_blank). Рассмотрите число кластеров K = 2, 5, 10, 20.

b) [DBSCAN](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html)

c) [AgglomerativeClustering](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html" \t "_blank). Рассмотрите число кластеров K = 2, 5, 10, 20.

1. Для каждой модели оцените потери от уменьшения цветов при помощи метрики SSIM.
2. Сформулируйте выводы по проделанной работе.
   1. Какая модель показала лучший результат?

Для получения зачета по этому домашнему заданию минимально должно быть реализовано обучение трех моделей с одним и тем же значением кластеров, где такой параметр присутствует.

**Результат**:

* Получены знания по применению моделей кластеризации.
* Форма выполнения:
* ссылка на Jupyter Notebook, загруженный на GitHub;
* ссылка на Google Colab;
* файл с расширением .ipynb.
* **Инструменты**:
* Jupyter Notebook/Google Colab;
* GitHub;
* палитра изображения.

Срок выполнения: дедлайн приема решений на проверку (18.09.2022)

\*Рекомендации к выполнению:

* Текст оформляйте в отдельной ячейке Jupyter Notebook/Google Colab в формате markdown.
* У графиков должен быть заголовок, подписи осей, легенда (опционально). Делайте графики бОльшего размера, чем стандартный вывод, чтобы увеличить читабельность.
* Убедитесь, что по ссылкам есть доступ на чтение/просмотр.
* Убедитесь, что все ячейки в работе выполнены и можно увидеть их вывод без повторного запуска.

Список источников

1. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
2. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>
3. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>
4. Пример выполнения задания <https://colab.research.google.com/gist/Irina-64/0ab39d5745e1dc5bbe62a4ad2116c2c3/trubchik_homework-2-palitra-clustering.ipynb>