

Решение задачи кластеризации изображений картин художников с помощью программного пакета визуального программирования Orange

Голубева Евгения Павловна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Цель данной статьи – решить задачу кластеризации изображений картин художников. Для кластеризации изображений был использован программный пакет визуального программирования на основе компонентов для визуализации данных Orange и изображения картин художников Эдуарда Мане, Клод Моне и Эдварда Мунк. С помощью средств визуализации Orange решили задачу кластеризации изображений картин художников и получили итоговую схему.

Ключевые слова: Orange, виджет, изображения, кластеризация.

Solving the problem of clustering images of artists' paintings using the Orange visual programming software package

Golubeva Evgeniya Pavlovna

Sholom-Aleichem Priamursky State University

Student

Abstract

The purpose of this article is to solve the problem of clustering images of paintings by artists. To cluster images, a visual programming software package based on Orange data visualization components and images of paintings by artists Edouard Manet, Claude Monet and Edvard Munch was used. With the help of Orange visualization tools, we solved the problem of clustering images of artists' paintings and obtained the final scheme.

Keywords: Orange, widget, images, clustering.

1 Введение

1.1 Актуальность

В последние годы наблюдается значительный рост интереса к применению искусственного интеллекта и машинного обучения в различных областях, включая искусство. Кластеризация изображений является одним из ключевых приложений этих технологий.

Кластеризация изображений картин художников позволяет автоматизировать процесс анализа и классификации художественных стилей,

что может быть полезно для искусствоведов, историков искусства и коллекционеров.

Orange — это мощный и удобный программный пакет для визуального программирования, который специализируется на анализе данных и машинном обучении. Он предоставляет пользователям интуитивно понятный интерфейс, позволяющий создавать и анализировать данные без необходимости написания кода. Orange поддерживает широкий спектр методов и алгоритмов, включая кластеризацию, классификацию, регрессию и визуализацию данных.

1.2 Обзор исследований

Н.Ю. Ильясова, А.В. Устинов, А.Г. Храмов проводили экспериментальные результаты кластеризации тестовых и натуральных изображений лейкоцитов крови [1]. Рассматривали этапы решения задач кластеризации изображений с использованием предобученных нейронных сетей А.С. Кузнецов, Е.Ю. Семенов, Л.Д. Матросова [2]. А.С. Титов предложил методику диагностирования с применением алгоритмов кластеризации спутниковых снимков [3]. В статье рассмотрел использование методов кластеризации в программе Orange на основе реальной базы данных. Н. Юсупов [4]. А. В. Леонов в статье рассматривал основные алгоритмы кластеризации категориальных данных применительно к различным типам пользовательских интерфейсов, определяются их достоинства и недостатки [5].

1.3 Цель исследования

Цель исследования - решить задачу кластеризации изображений картин художников.

2 Материалы и методы

Для решения задачи кластеризации изображений картин художников используется программа Orange. Работа будет происходить на готовом наборе данных с изображениями картин художников Эдуарда Мане, Клод Моне и Эдварда Мунк скачать которые можно по ссылке:

https://drive.google.com/file/d/1SJloCYOoAmhxcWBqprF1al-ZDZ_3Dd7g/view?usp=sharing

3 Результаты и обсуждения

Перед началом работы требуется установить Orange с официального сайта и установить.

Создадим новый файл (см.рис.1).



Рисунок-1 Создание нового файла

Для решения задачи кластеризации изображения необходимо установить дополнение Image Analytics. Для того, чтобы скачать дополнение, необходимо перейти в Options, далее в Add-ons, в появившемся окне выбираем Image Analytics (см.рис.2).

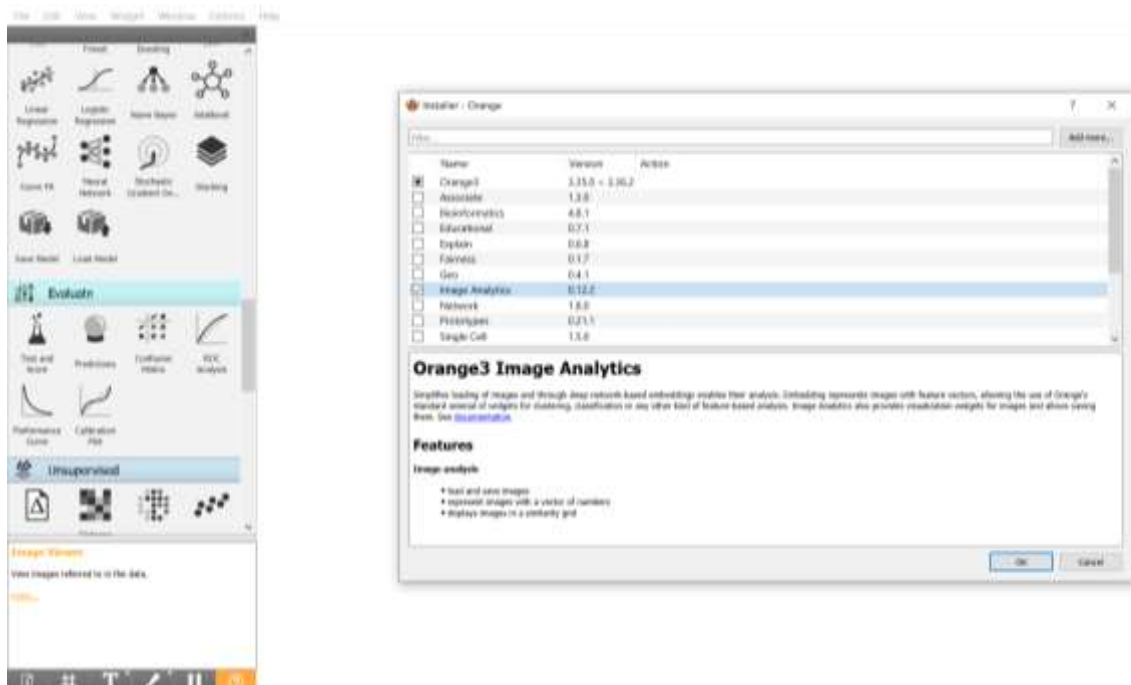


Рисунок-2 Установка дополнения Image Analytics

Для того, чтобы загрузить изображения, необходимо из раздела Image Analytics выбрать виджет Import Images и перенести его на холст (см.рис.3).



Рисунок- 3 Добавление виджета Import Images на холст

Открываем виджет Import Images и подгружаем подготовленную папку с изображениями картин художников Эдуарда Мане, Клод Моне и Эдварда Мунк. После того как подгрузили папку появилась информация, что папка с букетами цветов содержит 3 категории и 45 картинок (см.рис.4).



Рисунок- 4 Загрузка папки с изображениями букетов цветов

Для того чтобы посмотреть загруженные изображения добавляем виджет Image Viewer, и соединяем с виджетом Import Images (см.рис.5).



Рисунок 5- Добавление виджета Image Viewer

Открываем виджет Image Viewer. В появившемся окне можно увидеть изображения картин художников, а также и их названия (см.рис.6).



Рисунок- 6 Просмотр изображений

Далее добавим виджет Image Embedding на холст и соединим с виджетом Import Images (см.рис.7). Виджет Image Embedding получает таблицу с изображениями и отправляет их на сервер для встраивания в формате, понятном алгоритмам машинного обучения. Сервер проталкивает изображения через предварительно обученную глубокую нейронную сеть и возвращает числовые векторы в виджет.

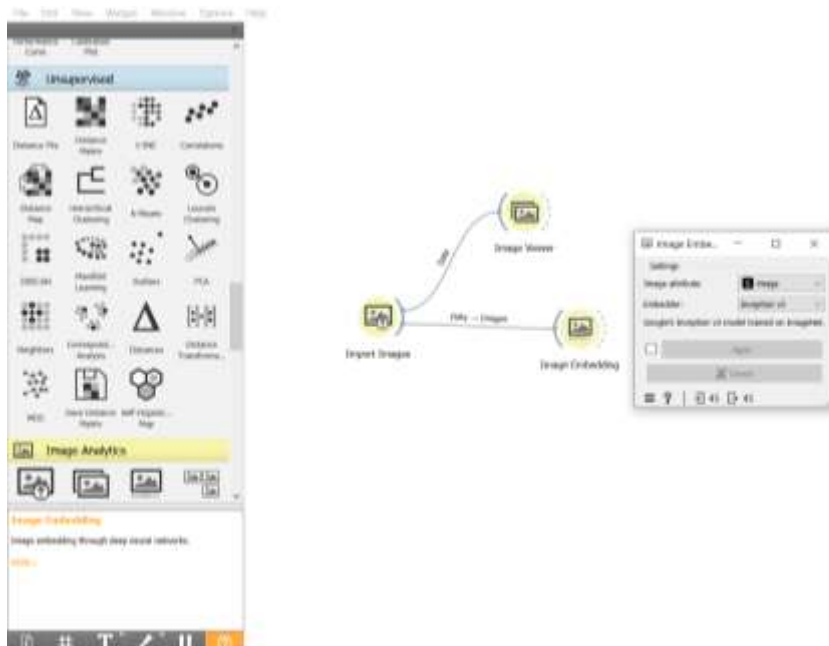


Рисунок- 7 Добавление виджета Image Embedding

Для того чтобы просмотреть данные виджета Image Embedding, добавляем виджет Data Table и соединяем с Image Embedding (см.рис.8).

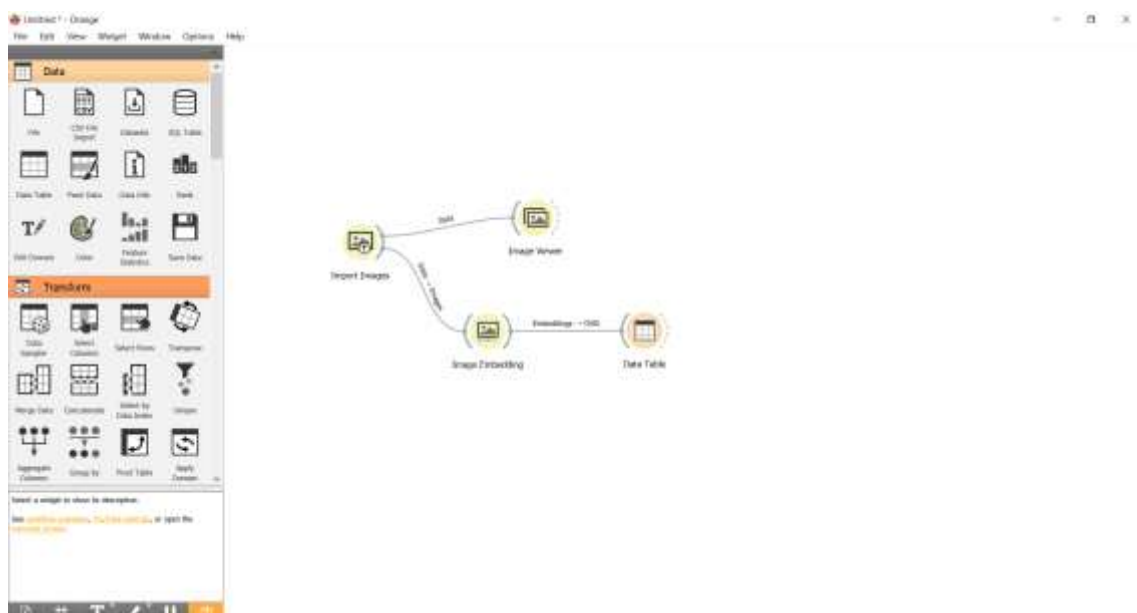


Рисунок- 8 Добавление виджета Data Table

Открываем виджет Data Table. С помощью таблицы можем увидеть, что для каждого изображения добавилось 2048 дополнительных признаков (см.рис.9).

Image name	size	width	height	...
A Bar of the Fool	252344	5334	354	0.113429
Apprentice	279376	196	361	0.488253
Battle Hound	302773	3371	3332	0.333586
Call-Garden	255691	371	381	0.277388
In the Cosmos	412171	1280	875	0.306370
In Heaven as in Hell	31942	1331	1334	1.28881
Just from the Past	256695	3234	1319	0.618177
Music in the Sub	212919	982	1938	1.02143
Street of the	244523	2311	2304	0.262821
Street of the	191726	380	399	0.407754
The Archer	1741864	3282	2362	0.178288
The Bakery	164648	160	398	0.255919
The girl in the s...	19373	233	300	0.252443
The Spanish G...	47595	465	515	0.934424
Edward Manet	4659	491	453	0.354708
Sea of the S...	239753	2776	2242	0.628668
Carle Monet	291137	1985	1447	1.01103
Impressionist	28417	399	324	1.29189
In Island Card	8812919	2282	1198	0.176247
Monet's garden	241164	1470	1138	0.255241
Monet on the	46073	1280	1328	0.402577
Paul Signac	138838	1440	1289	0.222526
The City at Br...	103375	1480	1278	0.223403
The House of ...	112979	2149	1918	0.287498
The Marine	275052	4339	2809	0.22921
Trip from the H...	163708	2168	2844	0.417543
Water Lilies (c...	303986	6231	1803	0.228487
Water Lilies	17571	441	408	0.344874
Water Lilies	35263	380	357	0.475081
Widow Weepin...	24267	1534	1429	0.184483
Widow	36389	539	530	0.348879
Widow	13801	331	338	0.441191
Widow	21894	335	332	0.337782
Widow	21894	713	332	0.46796
Widow	17898	343	332	0.23022
Widow	17919	380	419	1.02667

Рисунок-9 Просмотр данных виджета Image Embedding

Следующим шагом добавляем виджет Distances на рабочую область и устанавливаем связь с виджетом Image Embedding. Затем открываем настройки виджета Distances и выбираем метрику расстояния cosine (см. рис. 10).

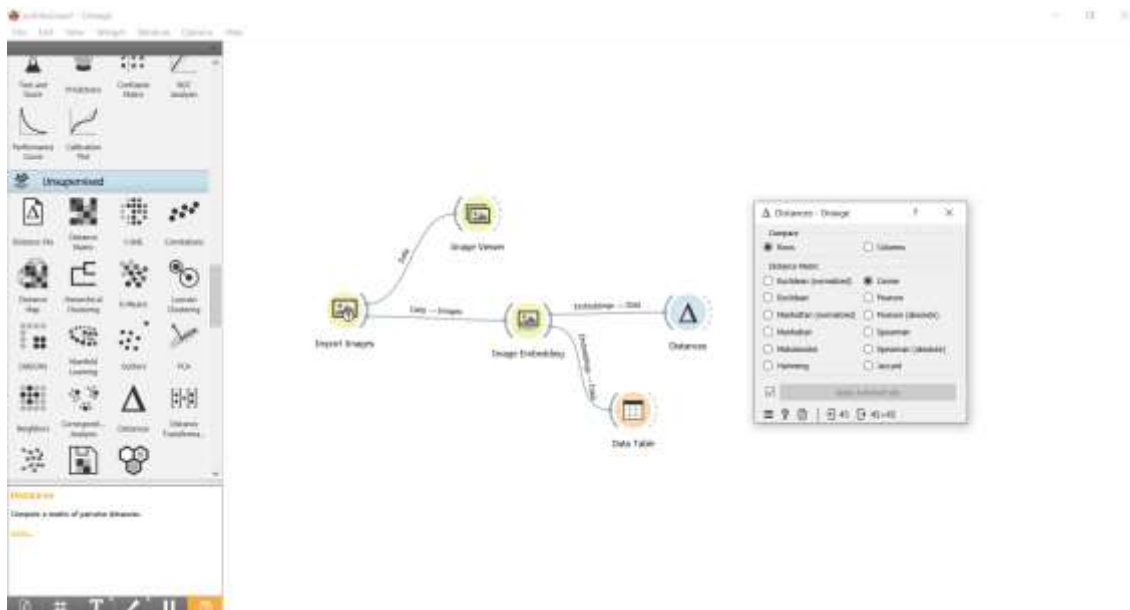


Рисунок -10 Добавление виджета Distances на холст

Далее выполним кластеризацию изображений. Для этого помещаем виджет Hierarchical Clustering на холст и создаем связь с виджетом Distances (см. рис. 11).

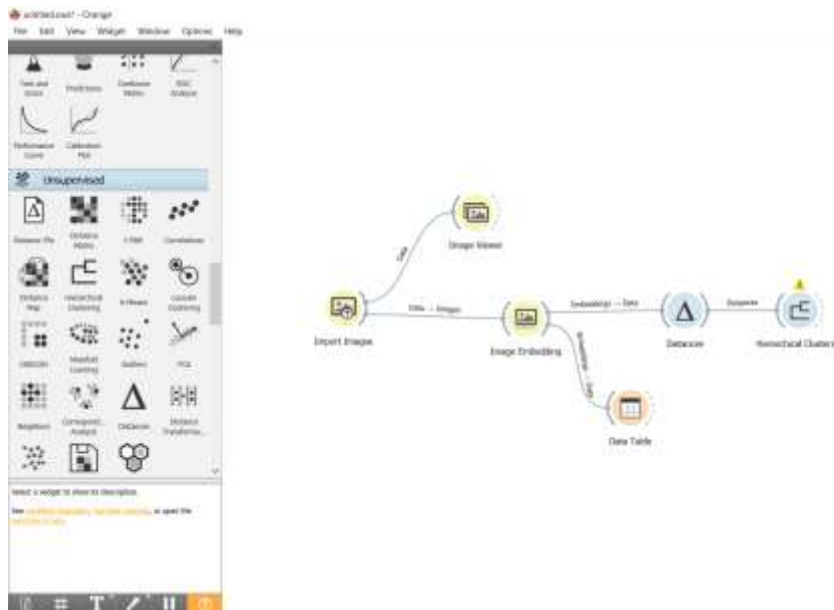


Рисунок -11 Добавление виджета Hierarchical Clustering на холст

Открываем виджет Hierarchical Clustering. В появившемся окне настраиваем условия иерархичной кластеризации. Для этого выбираем в расстоянии между кластерами (Linkage) связь Уорда, которая вычисляет увеличение суммы квадратов ошибок. Другими словами, критерий минимальной дисперсии Уорда минимизирует общую дисперсию внутри кластера. В окне обрезка (Pruning) выбираем максимальную глубину диаграммы 10. В окне выбор (Selection) выбираем количество количества верхних узлов 3. После настройки условий, можно увидеть, что образовались три группы, соответствующие художникам Эдуарда Мане, Клод Моне и Эдварда Мунк (см.рис.12).

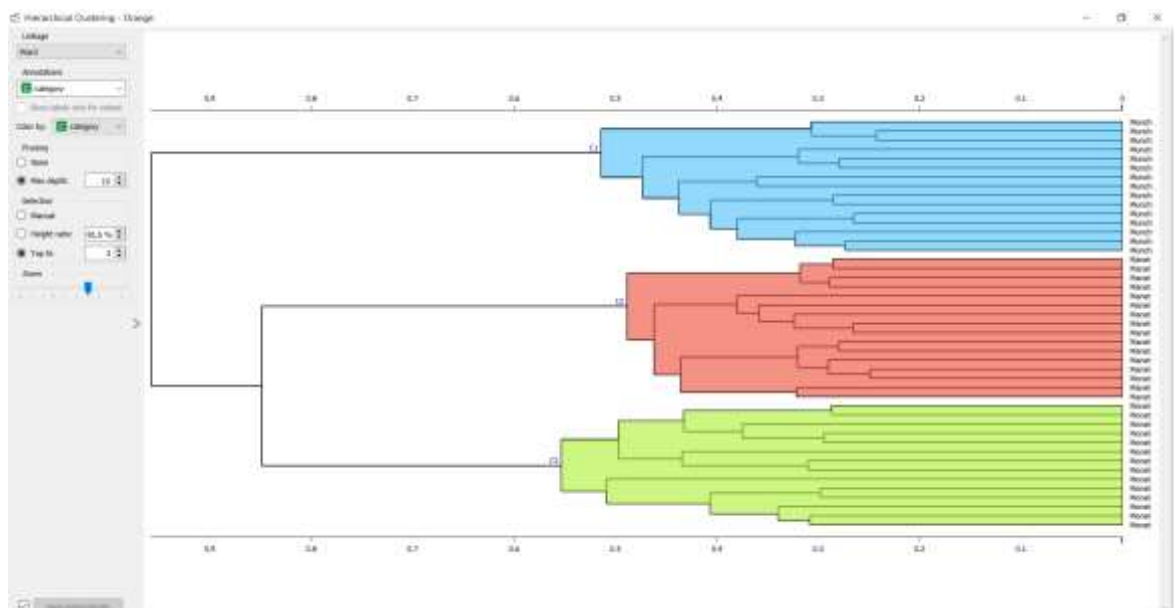


Рисунок - 12 Распределение изображений по кластерам

Также можем увидеть, что в группе С2 которая относится к художнику Эдуарду Мане, есть одна картина Клода Моне (см.рис.13).

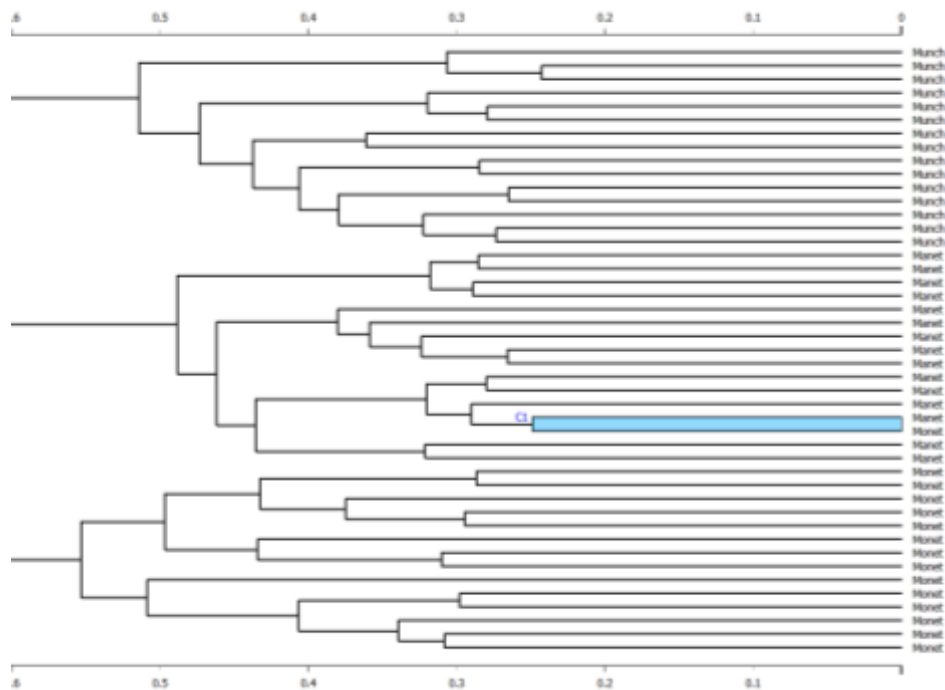


Рисунок -13 Ошибка к группе С2

Для того, чтобы посмотреть картину Клода Моне, которая отнеслась к группе картин Эдуарда Мане, добавляем виджет просмотра изображений Image Viewer и соединяем с виджетом Hierarchical Clustering (см.рис.14).

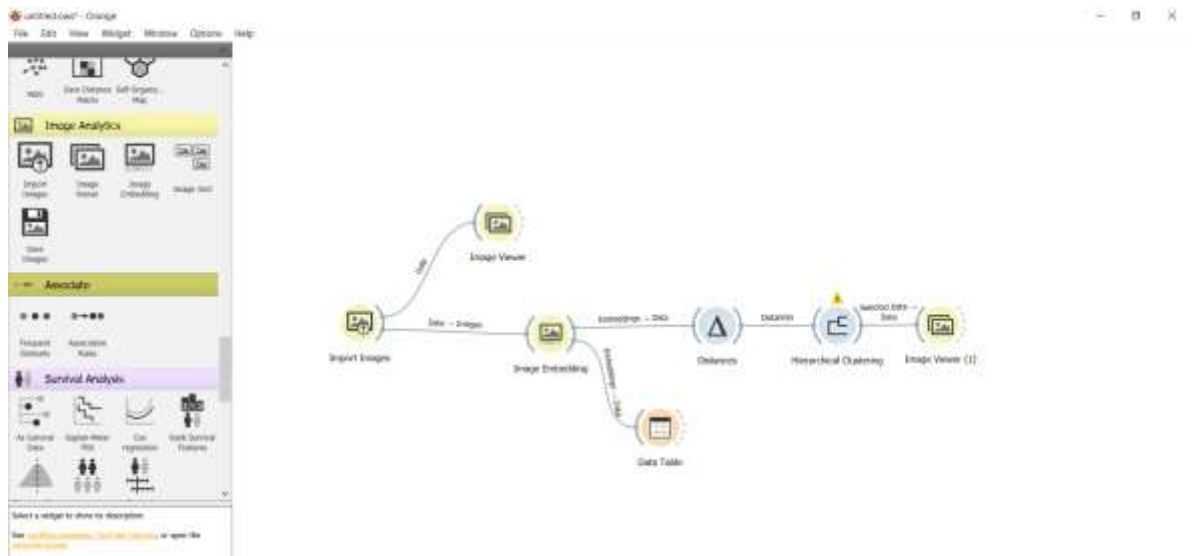


Рисунок 14- Добавление виджета Image Viewer

Открываем виджет Image Viewer. В Hierarchical Clustering выберем результат c1 (см.рис.), и изображения автоматически появятся в Image Viewer. В окне Image Viewer мы наблюдаем, что действительно результат c1 содержит изображения картины Клода Моне, которая очень схожа с

картиной Эдуарда Мане, поэтому алгоритм кластеризации ошибочно отнес её к группе С2. Этот пример подчеркивает важность выбора подходящих параметров и метрик для кластеризации, чтобы минимизировать такие ошибки (см.рис.15).

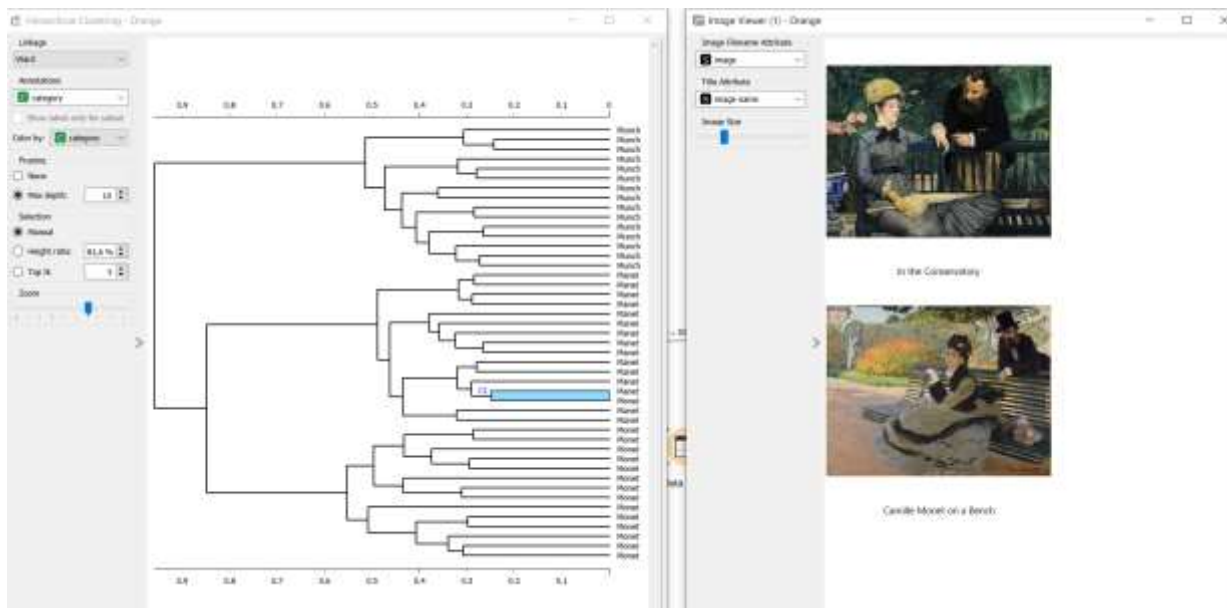


Рисунок-15 Проверка ошибочного результата

В итоге получилась готовая схема, с помощью которой можно решить задачу кластеризации изображений (см.рис.16).

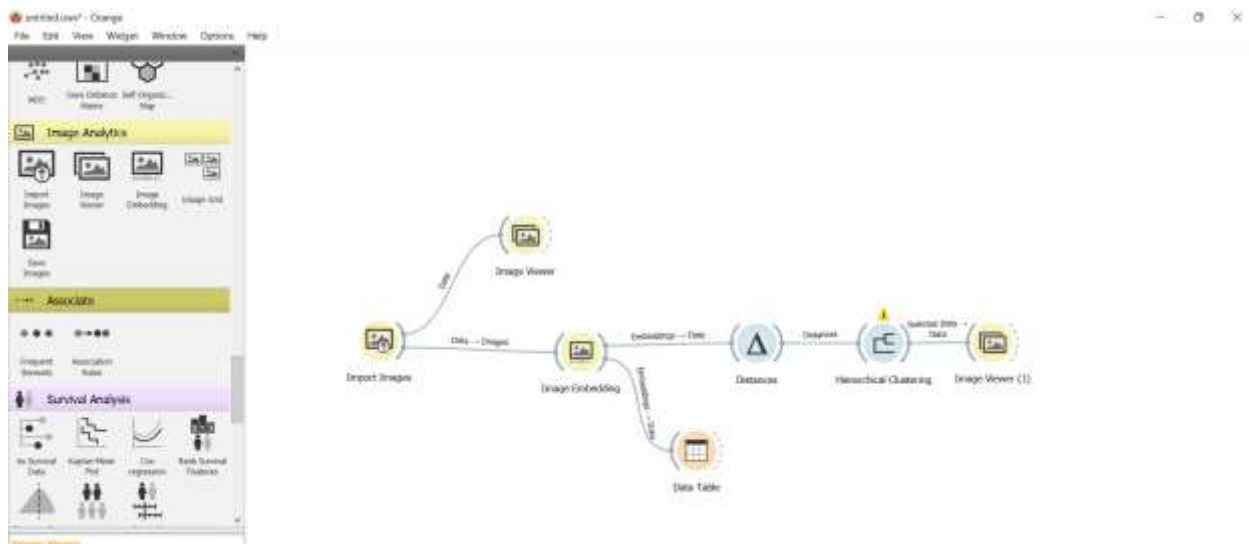


Рисунок-16 Итоговая схема

Выводы

В данной работе была выполнена задача кластеризации изображений картин художников Эдуарда Мане, Клод Моне и Эдварда Мунк. С помощью виджетов Import Image, Image Viewer, Image Embedding, Data Table, Distances,

Hierarchical Clustering выполнили кластеризацию изображений картин художников и получили итоговую схему.

Библиографический список

1. Кузнецов А. С., Семенов Е. Ю., Матросова Л. Д. Кластеризация изображений при использовании предобученных нейронных сетей // International journal of open information technologies. 2019. Т. 7. №. 7. С. 42-47.
2. Моисеенко Г. А. и др. Классификация и распознавание изображений живой и неживой природы //Оптический журнал. 2015. Т. 82. №. 10. С. 53-64.
3. Юсупов Н. Исследование методов кластеризации в программе Orange //Молодежная школа-семинар по проблемам управления в технических системах имени АА Вавилова. 2020. Т. 1. С. 35-37.
4. Клименко А. В., Слащев И. С. кластерный анализ данных //Вестник науки. 2019. Т. 1. №. 1. С. 159-163.
5. Гринченков Д. В. и др. Сравнительный анализ алгоритмов интеллектуального анализа данных //Моделирование. Теория, методы и средства. 2016. С. 263-266.
6. Мастевой С. С., Петрова А. Н. Data mining: обзор методов и области их применения // Наука, инновации и технологии: от идей к внедрению. 2022. С. 38-40.
7. Ильясова Н. Ю. и др. Алгоритмы автоматической кластеризации изображений микрочастиц //Компьютерная оптика. 1993. №. 13. С. 39-46.