Slide 1 : Меня зовут Ле Ву Бинь, сегодня я представлю тему "Автоматическое программирование и аннотирование сценариев обработки изображений в биоинформатике", под руководством доцент

Козлов Константин Николаевич

Slide 2 : В данной теме разработана система автоматической обработки изображений, которая поможет нам обрабатывать заданные графики обработки изображений. Эта система делает процесс обработки интуитивно понятным и высокоточным.

Алгоритмы обработки изображений встроены в модуль function.py с помощью таких инструментов, как Prostak , OpenCv ,

Область применения : Во всей области обработки изображений, но приоритет отдается биологии и процессу биологических исследований

Методы исследования - Разработка программ, тестирование, анализ, испытание, сравнение

Slide 3 : С непрерывным развитием биологии количество необработанных данных изображений растет с каждым днем. Эти изображения являются инструментом биологических открытий, что привело к необходимости анализа и обработки этого источника необработанных данных.

Slide 4 : Написать программу, которая будет по имеющемуся графу генерировать код на Python, который можно будет редактировать: добавлять или исключать операции, заменять операции, и т.д.,

Вместе с кодом должно создаваться текстовое описание процедуры, в котором описываются все шаги и основные свойства операций

В данной теме представлен подход к решению указанной проблемы: он заключается в построении программы автоматической обработки изображений, объединяющей технологии обработки изображений

Slide 5 : Проект реализуется в основном с помощью языков Python и R. Два языка были выбраны из-за сильных сторон, которые они привносят в таблицу. Хотя у него есть свои слабые стороны, они не слишком сильно влияют на мою диссертацию. Кроме того, в python поддерживается множество пакетов конфигурации для обработки изображений, а язык R помогает нам сделать статистику данных, полученных после обработки, более наглядной.

Slide 6 :

Система использует множество методов анализа и обработки изображений, среди которых можно выделить два очень важных метода, таких как метод Канни и пороговый метод.

Фильтр Кэнни представляет собой комбинацию множества различных шагов для поиска и оптимизации границы, в результате чего получается довольно тонкая и точная граница. Процесс поиска ребер с использованием метода Канни можно выполнить, выполнив следующие 4 шага:

1. Удалите шум на изображениях

2. Рассчитайте значение градиента на изобр

3. Исключите немаксимальные значения

4. Выберите край объекта на изображении

Для каждого пикселя применяется одно и то же пороговое значение. Если значение пикселя меньше порогового значения, оно устанавливается равным 0, в противном случае оно устанавливается на максимальное значение. Функция cv.threshold используется для применения порога. Первый аргумент — это исходное изображение, которое должно быть изображением в градациях серого. Второй аргумент — это пороговое значение, которое используется для классификации значений пикселей. Третий аргумент — это максимальное значение, которое присваивается значениям пикселей, превышающим пороговое значение

Вопрос : В процессе обработки изображений в биологии какой метод является наиболее важным?

Slide 7 :

В ходе реализации проекта я консультировался и обращался за помощью к Prostak и OpenCv для создания полного набора данных о методах обработки изображений.

Prostak

Программа обработки изображений ProStack, в которой последовательность операций задается визуально в виде графика. Описания доступных процедур хранятся в базе данных SQLite. Архитектура ProStack открытая, но требует доработки и отладки для подключения дополнительных программ. Граф сценария сохраняется в текстовом файле INI.

OpenCv

OpenCV позволяет увеличить скорость вашего процессора при выполнении операций в реальном времени. Он также предоставляет большое количество кода для обработки методов обработки изображений.

Slide 8 :

Теперь перейдем к основному алгоритму работы автоматической системы:

Сначала создаю класс Node:Node, который представляет каждый этап обра-  
ботки графа и хранит адрес входного изображения и адрес выходного изображения, номер узла в графе и переменную, которая проверяет, завершен ли вход Это завершено или нет?

Читаю блок схема !

Можно вопрос здесь есть ?

Slide 9 :

Давайте рассмотрим конкретные примеры, чтобы лучше понять систему и ее эффективность

Есть изображения глаза развивающегося дрозофилы.

Там 3 группы:

1. female

2. male

3. pseudomale

Использовано 5 генов + маркер ядра DAPI:

1 DI - delta

2 dsx – doublesex

3 Hh – hedgehog

4 Sxl- Sex-lethal

5 ms2 - Male-specific 2

В изображениях яркие точки – комплексы молекул РНК, которые надо найти – сегментировать. Также требуется определить границы клеток. Для этой задачи есть сценарии.

Slide 10 :

Мы можем видеть иллюстрации 5 генов + маркер ядра DAPI так :

Slide 11 :

Однако все найденные точки считались равноправными, т.е. содержащими одинаковое количество молекул РНК. Это можно скорректировать ипользуя гистограмму суммарной интенсивности этих точек. Там может быть несколько пиков, самый большой будет соответствовать простейшему случаю – одна молекула=одна точка. Пики ниже и правее могут соответствовать “слипшимся” молекула, т.е. эти точки надо считать как две молекулы.

Мы получаем график суммы величин этих точек следующим образом :

Slide 12 :

После построения гистограммы мы найдем вершины графика и определим каждый диапазон значений для каждой из этих вершин, первый интервал соответствует 1 молекуле, второй интервал соответствует 2 молекулам, продолжайте так

У нас еще есть графиков количества молекул в полосе по х от 450 до 550 следующим образом :

Slide 13 :

Текущая обработка изображений крайне необходима. Это помогает в изучении врожденных проблем или в области медицины. В этом проекте я создал простую автоматическую систему обработки биологических изображений в порядке заданного графа. Эта система включает в себя:

1. Создаваемая разработанным методом программа обработки изображений позволяет гибко настраивать параметры элементарных процедур.

2. Автоматическое создание текстового описания сценария обработки облегчает совместную разработку таких сценариев.

3. Анализ распределения суммарной интенсивности в сегментированных объектах указывает на наличие парных и тройных комплексов.

Таким образом, простая система была завершена, но эта система удовлетворительна только с точки зрения обработки изображений, но с ограниченным объемом данных, а в реальности нам часто приходится обрабатывать бесконечное количество данных. Поэтому, чтобы продолжить и завершить эту тему, нам необходимо разработать больше вспомогательных систем для обработки больших данных.Это будет следующим направлением и развитием, чтобы сделать тему все более и более приближенной к реальности и полезной.