

Санкт–Петербургский государственный университет

ТРУШОВ Константин Андреевич

Выпускная квалификационная работа
Мониторинг сельскохозяйственных площадей по
спутниковым снимкам

Уровень образования: магистратура

Направление 02.04.02 «Фундаментальная информатика и информационные
технологии»

Основная образовательная программа ВМ.5503.2021
«Технологии баз данных»

Научный руководитель:

доцент, кафедра компьютерного моделирования
и многопроцессорных систем,
к.т.н. Гришкин Валерий Михайлович

Рецензент:

старший инженер-программист,
ООО «Ядро Центр Технологий
Мобильной Связи»,
к.т.н. Епифанов Николай Анатольевич

Санкт-Петербург

2023 г.

Содержание

Введение	3
Постановка задачи	4
Обзор литературы	5
Глава 1. Сбор и предобработка тренировочной выборки	6
1.1. Начальный набор данных	6
1.2. Предобработка набора данных	8
1.3. Влияние различных подмножеств частот на обучение модели	10
Глава 2. Обзор архитектур использованных моделей	12
2.1. UNet	12
2.2. UNet3+	14
2.3. Архитектуры ViT	14
2.3.1 SwinUNet	15
2.3.2 TransUNet	17
Глава 3. Детали реализации	19
3.1. Оценка качества сегментации	19
3.2. Аугментация данных во время обучения	20
3.2.1 Пространственные аугментации	20
3.2.2 Сглаживание целевых значений	20
3.2.3 Стохастическая глубина	21
3.3. Анализ экспериментов	23
3.4. Индекс вегетации	24
3.5. Итоговая программа	26
Заключение	28
Список литературы	29
Приложение 1А. Подробные метрики моделей SwinUNet и TransUNet	31
Приложение 1Б. Подробные метрики модели UNet3+	32

Введение

Точная классификация культур представляет собой сложную задачу, связанную, во-первых, с высокой внутрикласовой спектральной изменчивостью отдельных культур в течение вегетационного периода (фенологическое развитие), во-вторых, с высоким межклассовым спектральным сходством типов культур, и, в-третьих, в неизбежном наличии облаков на изображениях.

При использовании спутниковых снимков, снятых с помощью оптического прибора, в этой работе проводится исследование попиксельной классификации (сегментации) сельскохозяйственных культур, растущих на участках ферм. В частности, в исследовании рассматриваются результаты классификации культур за счет использования разных сочетаний диапазонов длин волн.

Наработки из данного исследования пригодятся в реализации системы, задача которой - сегментировать поля, определять тип растущей на них культуры и подсчитывать степень вегетации. Клиенты такой системы смогут удалённо отслеживать состояние своих участков и принимать решения на основе этой информации.

Постановка задачи

Целью работы ставится разработка методов, способных предоставить информацию о сельскохозяйственном участке по данным со спутниковых снимков. Для достижения поставленной цели потребовалось выполнение следующих задач:

- Найти данные, привести их к виду, готовому к обучению моделей;
- Рассмотреть актуальные архитектуры, решающие задачу сегментации, обучить их;
- Сравнить результаты обучения;
- Рассмотреть методы вычисления степени вегетации и выбрать наиболее подходящий;
- Обернуть разработанные методы в программу.

Обзор литературы

Основные информационные источники, использованные при написании работы:

- Литература по архитектурам использованных нейронных сетей: [1], [2], [12], [4], [5], [13];
- Информация о спутниковых приборах: [6], [3], [10];
- Литература по аугментациям данных: [15], [16];
- Информация о датасете: [8].

Глава 1. Сбор и предобработка тренировочной выборки

1.1 Начальный набор данных

Sentinel-2 — два спутника дистанционного зондирования Земли, запущенных в 2015 году. Производят оптические снимки. Делают это в 12 волновых диапазонах[6]. Снимок одного и того же участка могут производить с периодом в 5 дней. Пролетают через все континенты, прибрежные зоны и внутренние моря. Снимки можно получить на Copernicus Open Access Hub[7].

Номер	Название диапазона	Центральная длина волны, нм	
		Sentinel-2A	Sentinel-2B
1	Прибрежные воды/Аэрозоль	442.7	442.2
2	Синий	492.4	492.1
3	Зелёный	559.8	559.0
4	Красный	664.6	664.9
5	Красная граница растительности	704.1	703.8
6	Красная граница растительности	740.5	739.1
7	Красная граница растительности	782.8	779.7
8	NIR	832.8	832.9
9	Водяной пар	945.1	943.2
10	SWIR – Cirrus	1373.5	1376.9
11	SWIR	1613.7	1610.4
12	SWIR	2202.4	2185.7

Таблица 1: Волновые диапазоны спутников группы спутников Sentinel-2

На первых этапах обучение моделей производилось с помощью датасета со спутниковыми изображениями некоторого региона с фермами в Бранден-

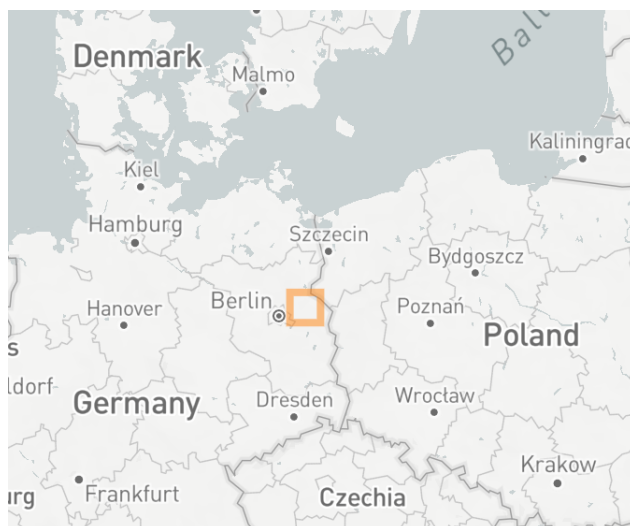


Рис. 1: Участок, размеченный в датасете

бурге, Германия[8]. Участок, снимки которого находятся в данном наборе данных, можно увидеть на рис. 1. Часть этого набора данных содержит полигоны полей в формате GeoJSON[9] с несколькими классами сельскохозяйственных культур, а также оптические снимки с группы Sentinel-2. Снимки данного региона присутствуют за 2018 и 2019 годы, в очищенную выборку вошли 144 снимка всей области из разных моментов времени. Дополнительно в датасете содержатся снимки в ещё двух дополнительных экземплярах, сняты они другими спутниковыми системами: Planet Fusion и Sentinel-1, но в обучении не использовались.

В представленном наборе данных были следующие 9 сельскохозяйственных культур:

- Wheat (Пшеница);
- Rye(Рожь);
- Barley (Ячмень);
- Oats (Овёс);
- Corn (Кукуруза);
- Oil Seeds (Масличные культуры);
- Root Crops (Корнеплоды);

- Meadows (Луга);
- Forage Crops(Кормовые культуры).

Виды культур были сильно разбалансированы между собой (рис. 2), в связи с чем было необходимо взвешивать ошибку для каждого класса при обучении моделей: умножать значение ошибки на коэффициент, обратный доле класса в выборке. Следовательно, классы, в меньшей степени представленные в наборе данных, имели больший вес, верно и обратное.

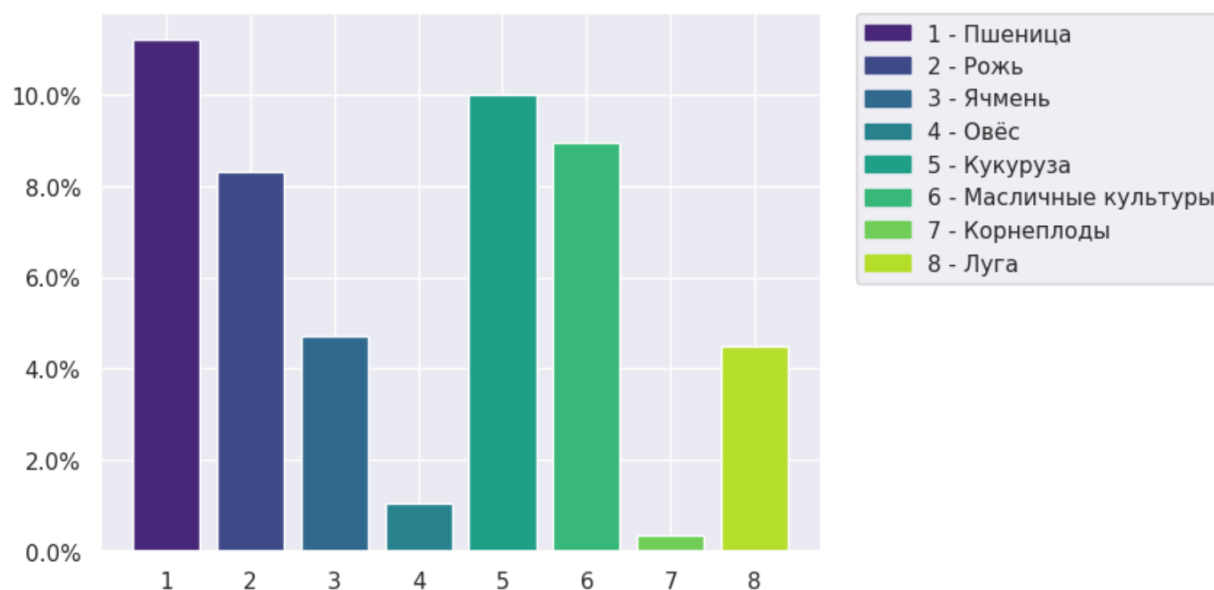


Рис. 2: Баланс классов в начальной выборке

1.2 Предобработка набора данных

Набор данных состоит из снимков разрешением 2400 на 2400 пикселей. В зависимости от входного слоя моделей, эти снимки разбивались на изображения поменьше: 256 на 256 или 224 на 224 пикселей.

К каждому снимку прилагается матрица с попиксельными вероятностями облаков. Данная матрица имеет меньшее пространственное разрешение, чем сам снимок. С помощью этой матрицы были отсеяны снимки, большую часть которых закрывают облака. Этот процесс происходил следующим образом: считалась средняя вероятность облаков на снимке, после чего, если эта

вероятность больше 0.4, снимок исключался из набора данных. Такой коэффициент был подобран эмпирическим путём, с помощью визуального осмотра отфильтрованной коллекции. Пример участка с наличием облака можно увидеть на рис. 3. В большинстве случаев на отсеянных снимках облако было видно во всех волновых диапазонах.

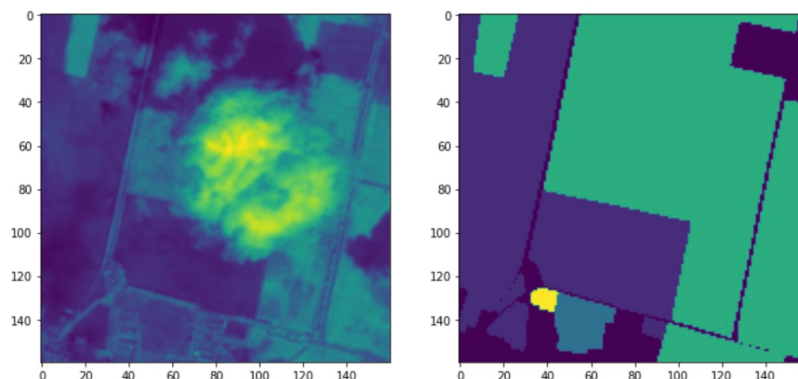


Рис. 3: Вхождение выборки с наличием облака

Перед обучением сети данные полигонов сельскохозяйственных полей были растеризованы так, чтобы их размеры соответствовали размерам снимков. В дополнение, участки, площадь которых меньше 1 гектара, были отсеяны, т.к. их размер слишком мал. Пример растеризованных полигонов можно увидеть в правой части рис. 4, где отдельные цвета обозначают отдельные классы агрокультуры.

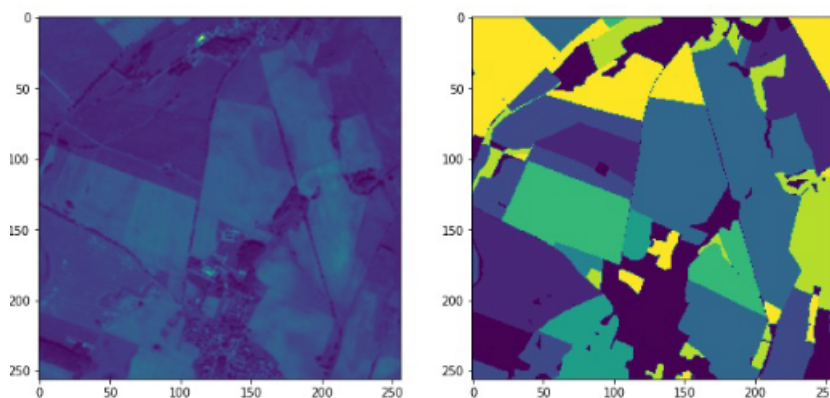


Рис. 4: Пример вхождения тренировочной выборки с большим количеством полигонов

Наборы диапазонов	precision	recall	F1
Полный	0.96	0.95	0.96
Agriculture	0.94	0.91	0.92
Color Infrared	0.92	0.89	0.91

Таблица 2: Средние значения метрик UNet на разных наборах

1.3 Влияние различных подмножеств частот на обучение модели

Было необходимо понять, нужно ли использовать все 12 частот или же будет достаточно всего нескольких из них. Для эксперимента были выделены следующие подмножества[3] волновых диапазонов:

- Color Infrared (№5, №4, №3) - ближний инфракрасный набор волновых диапазонов. Этот набор в данной задаче интересен тем, что его отражает хлорофил.
- Agriculture (№6, №5, №2) - агрикультурный набор. Из-за использования коротких и ближних инфракрасных волн здоровый урожай будет виден в зелёном цвете, а земля - в розоватом.
- Полный набор из 12-ти диапазонов.

На каждом подмножестве по отдельности была обучена модифицированная модель UNet (рис. 9).

	precision	recall	f1-score	support
Nothing	0.97	0.98	0.98	1642676
Wheat	0.97	0.96	0.96	282196
Rye	0.97	0.97	0.97	245957
Barley	0.97	0.97	0.97	127338
Oats	0.98	0.94	0.96	34296
Corn	0.97	0.96	0.96	306888
Oil Seeds	0.97	0.97	0.97	276479
Root Crops	0.99	0.91	0.95	6561
Meadows	0.90	0.90	0.90	130196
Forage Crops	0.95	0.91	0.93	93141
accuracy			0.97	3145728
macro avg	0.96	0.95	0.96	3145728
weighted avg	0.97	0.97	0.97	3145728

Рис. 5: Метрики UNet для полного набора

Из метрик видно, что лучше всего показала себя модель, обученная на полном наборе. Стоит отметить, что обучение этого этапа эксперимента длилось дольше всего, а также сам процесс предсказания требует больших ресурсов. На втором месте - агрикультурный набор. Color Infrared показал себя хуже всех. Средние значения метрик можно увидеть в таблице 2. Можно сделать следующий вывод: если ресурсы не важны, то стоит использовать полный набор диапазонов, иначе - агрикультурный.

На рис. 5, рис. 6 и рис. 7 можно увидеть покласовые метрики UNet на соответствующих наборах диапазонов.

	precision	recall	f1-score	support
Nothing	0.94	0.98	0.96	1480130
Wheat	0.93	0.94	0.94	316909
Rye	0.97	0.94	0.95	296667
Barley	0.97	0.87	0.92	149772
Oats	0.99	0.91	0.95	49807
Corn	0.94	0.95	0.95	345028
Oil Seeds	0.93	0.92	0.92	248551
Root Crops	0.86	0.94	0.89	4018
Meadows	0.92	0.77	0.84	127424
Forage Crops	0.95	0.86	0.90	127422
accuracy			0.94	3145728
macro avg	0.94	0.91	0.92	3145728
weighted avg	0.94	0.94	0.94	3145728

Рис. 6: Метрики UNet для агрикультурного набора

	precision	recall	f1-score	support
Nothing	0.97	0.96	0.96	1619541
Wheat	0.81	0.98	0.89	271977
Rye	0.96	0.89	0.93	185693
Barley	0.93	0.94	0.94	148016
Oats	0.89	0.84	0.87	24041
Corn	0.95	0.96	0.95	360462
Oil Seeds	0.96	0.96	0.96	258907
Root Crops	0.85	0.75	0.80	3889
Meadows	0.94	0.74	0.83	119604
Forage Crops	0.97	0.91	0.94	153598
accuracy			0.94	3145728
macro avg	0.92	0.89	0.91	3145728
weighted avg	0.95	0.94	0.94	3145728

Рис. 7: Метрики UNet для набора Color Infrared

Глава 2. Обзор архитектур использованных моделей

2.1 UNet

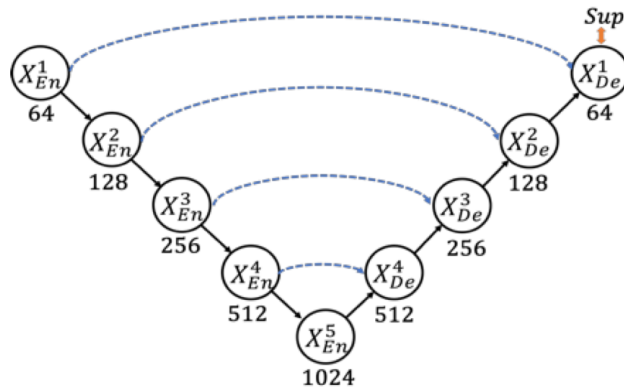


Рис. 8: Архитектура сети U-Net

Это свёрточная нейронная сеть, которая изначально была представлена для решения задач сегментации на медицинских изображениях, полученных при помощи микроскопа. В дальнейшем сфера использования этой архитектуры в задачах сегментации всё расширялась и расширялась.

Схема архитектуры представлена на рис. 8. Сеть состоит из ветки анализа, выделяющей признаки из входного тензора, и ветки синтеза, разворачивающей эти признаки в маску с размером немного меньшим, чем у входного изображения. Из ветки анализа в ветку синтеза идут несколько обходных

связей: таки образом, верхнеуровневые признаки соседствуют с нижеуровневыми, делая сегментацию более точной. В конце ветки синтеза есть слой, уменьшающий количество каналов до нужного нам, распределяя признаки по классам. Ветки состоят из нескольких блоков, которые в свою очередь содержат 2 последовательных слоя свёртки и один слой взятия максимумов (max pooling, в ветке анализа) или повышения размерности (upsampling, в ветке синтеза).

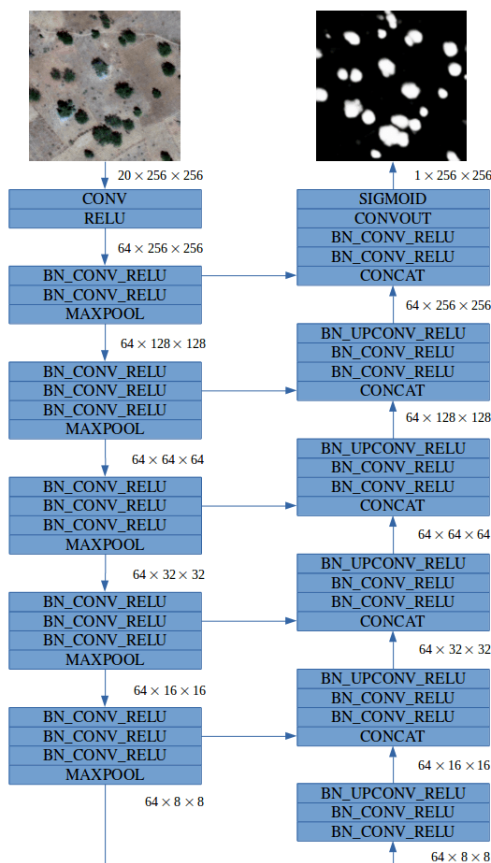


Рис. 9: Архитектура модифицированной сети UNet

Как можно увидеть на рис. 9, количество каналов во всех слоях нейронной сети равно 64, что очень мало по сравнению с оригинальной архитектурой. Этот шаг сделан исходя из следующих соображений:

- Можно позволить потерять немного информации за счёт обходных связей: ветка синтеза будет иметь доступ к низкоуровневым признакам.;
- Спутниковые изображения в имеемой выборке - плоские, т.е. снимки

представлены в двумерном виде, поэтому нет мотивации иметь в блоках модели много каналов с признаками.

2.2 UNet3+

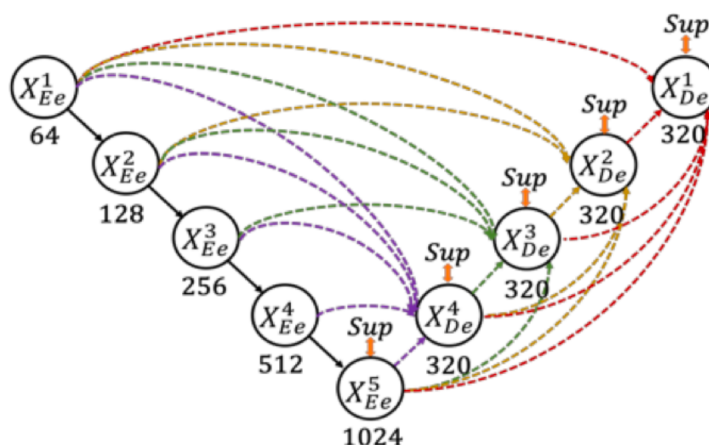


Рис. 10: Архитектура сети UNet3+

Свёрточная нейронная сеть на рис.10 использовалась авторами для решения задачи объёмной сегментации сканов компьютерной томографии[12]. Её особенность по сравнению с классической моделью (рис.8) в том, что связи между слоями энкодера и декодера находятся не только на одном уровне. Сеть так же состоит из ветки энкодера, выделяющей признаки из входного тензора, и ветки декодера, разворачивающей эти признаки в маску с размером немного меньшим, чем у входного изображения.

Принцип построения связей для ветки кодера: каждый слой кодера отправляет данные слоям декодера не только на своём уровне, но и на всех нижележащих. Принцип построения связей для ветки декодера: каждый слой декодера отправляет данные всем вышележащим слоям декодера. Авторы называют такой тип связей «full-scale skip connections». Таким образом, сеть на каждом этапе декодинга учитывает примитивные и верхнеуровневые признаки сразу вместе, поэтому качество предсказаний ожидаются более точными.

2.3 Архитектуры ViT

До недавнего времени все State-of-The-Art модели в задаче сегментации были исключительно свёрточными нейронными сетями. Однако всё измени-

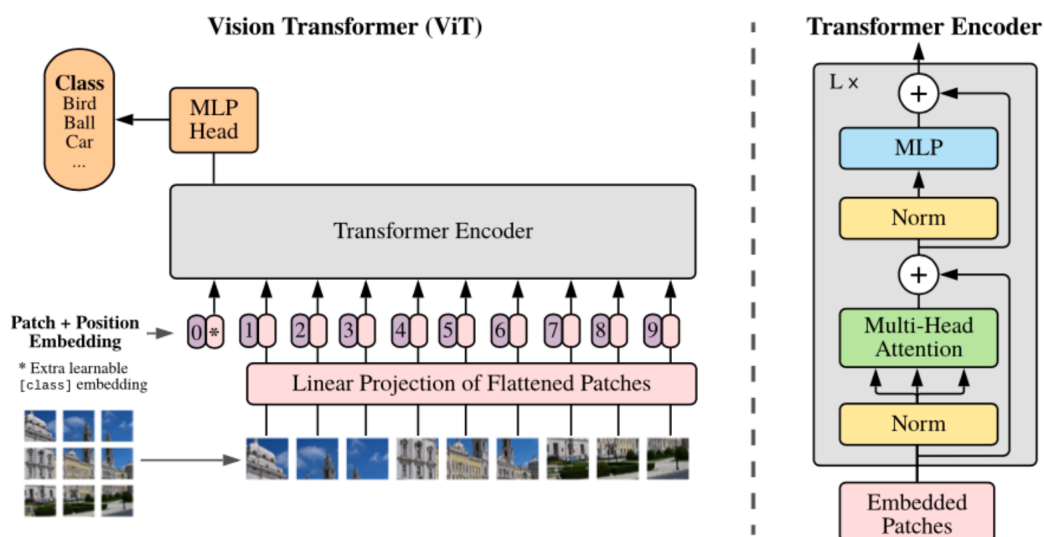


Рис. 11: Схема обработки изображения с помощью трансформера

лось, когда были предприняты первые попытки [11] использовать архитектуру Transformer из задач обработки естественного языка.

Тогда стоял вопрос, как именно представить изображение в токенизированном виде. Авторы оригинальной статьи предложили разбивать изображение на несколько частей (в статье - 16 на 16), каждую из них проецировать в скрытое представление (производить embedding), а над получившимся произвести позиционное кодирование (по аналогии с позицией слова в предложении). После чего задача получала тривиальный вид для исследователей обработки естественного языка.

Архитектуры свёрточных нейронных сетей, использующих Transformer достигают больших показателей по метрикам, чем обычные CNN, однако требуют предобучения на немалых объёмах данных.

Рассмотренные ниже архитектуры можно отнести к семейству ViT.

2.3.1 SwinUNet

Приняв во внимание последние исследования по ViT, Авторы SwinUNet [4] решили полностью отказаться от свёрточных блоков, заменив их на трансформеры. Как можно видеть на рис. 12, структурно архитектура похожа на классическую.

Идея архитектуры состоит в том, чтобы не терять мелкие детали изображения на самых первых слоях сети за счёт механизма self-attention. Здесь

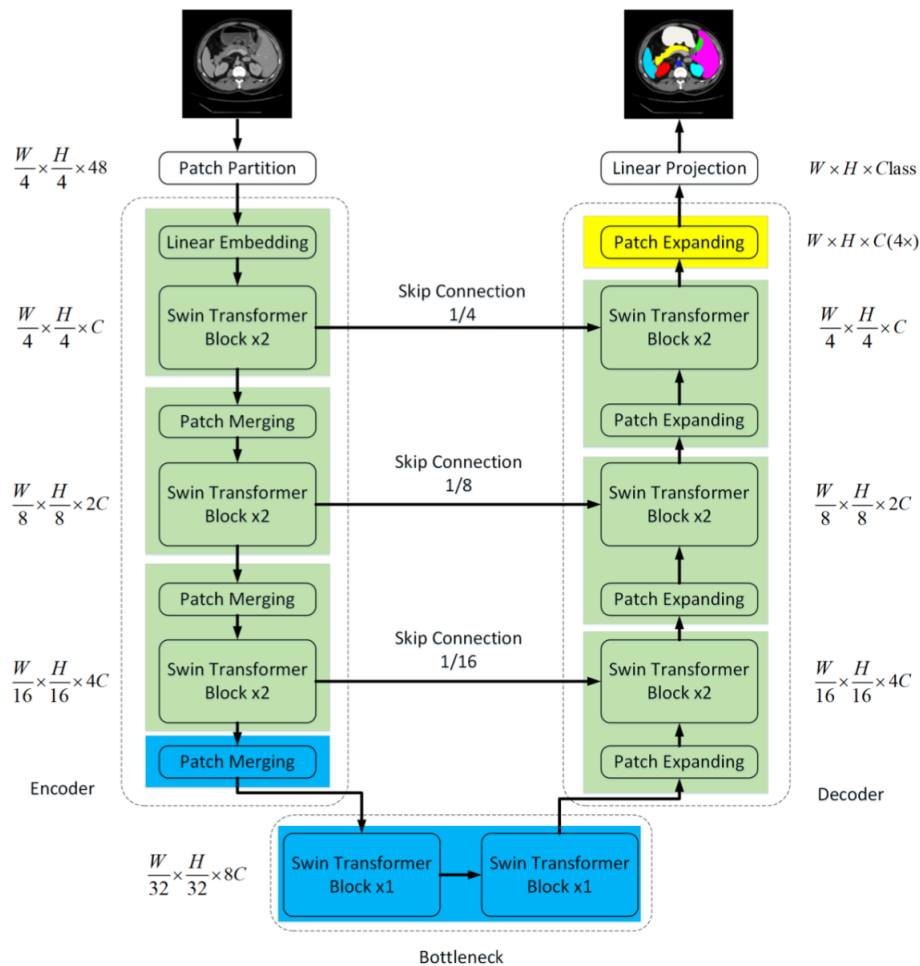


Рис. 12: Архитектура SwinUNet

изображение разрезается не на 16×16 частей, как в оригинальном ViT, а на 4×4 , чтобы можно было обработать более мелкий контекст.

Энкодер и декодер состоят из блоков под названием SwinTransformer [5], в которых слои нормирования, слои внимания (self-attention) и многослойные перцептроны взаимодействуют так, как показано на рис. 13. В целом механизм работы аналогичен классическому блоку трансформера, за исключением механизма внимания. Авторы предлагают в двух последовательных блоках-трансформерах использовать два немного отличающихся механизма внимания (рис. 14):

- В первом сначала идёт разделение на окна - патчи (небольшие участки изображения после обработки) группируются, далее на каждом окне отдельно вычисляется self-attention;
- Во втором также используются окна, но они сдвигаются так, чтобы

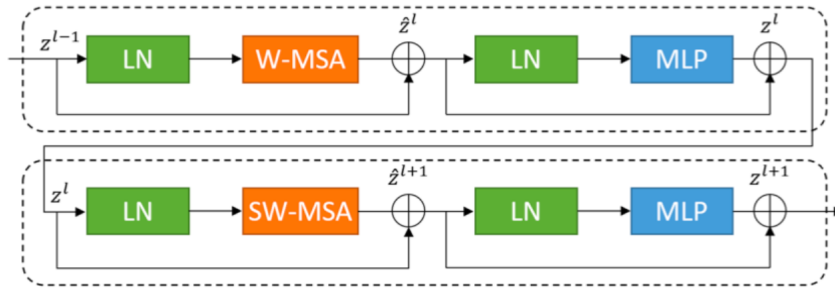


Рис. 13: Блоки SwinTransformer из SwinUNet[4]

пересекаться с окнами предыдущего шага;

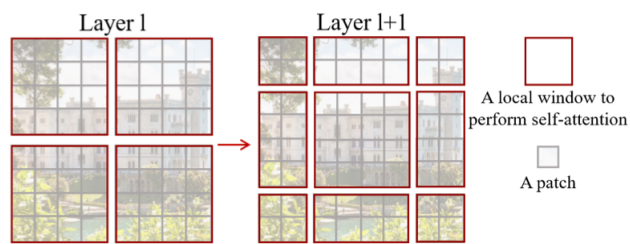


Рис. 14: Сдвиг окна в механизме внимания

Такой подход требует меньше времени на вычисления, чем альтернативный механизм со скользящим окном, показывая при этом схожие результаты.

2.3.2 TransUNet

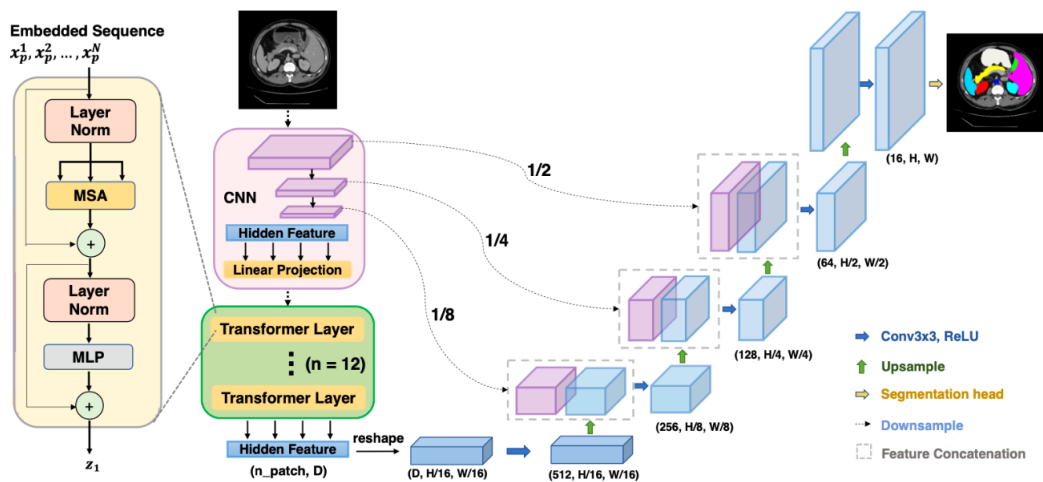


Рис. 15: Архитектура TransUNet

Данная архитектура отличается гибридным энкодером (рис. 15): сначала свёрточная сеть кодирует изображение, попутно передавая признаки в ветку декодера с помощью skip-connections. После чего выход свёрточной сети токенизируется и попадает в трансформер. Смысл такой архитектуры в том, что на этапе декодинга пространственные признаки из свёрточной сети и контекстуальные из трансформера обрабатываются совместно.

Глава 3. Детали реализации

3.1 Оценка качества сегментации

Одной из самых простых метрик является доля правильных ответов (1). Как следует из её названия, она показывает отношение количества правильных ответов к общему количеству ответов. Для итоговой оценки качества в задаче сегментации с множеством классов она не подходит, но тем не менее будет использована для отслеживания общего прогресса при обучении.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

К условиям текущей задачи запросто могут подойти метрики полноты (2) и точности (3). Полнота показывает, какую долю положительных предсказаний от общей доли положительных значений сделала модель. Точность показывает долю положительных предсказаний, которые являются таковыми в действительности.

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$precision = \frac{TP}{TP + FP} \quad (3)$$

В задачах сегментации часто используют коэффициент Dice (4), который в русской литературе может называться как коэффициент Сёренсена или Сёренсена-Дайса. Его показания идентичны F1-мере (5), которая по сути является балансом между полнотой и точностью. Метрика используется для работы с сегментированными областями. Она показывает отношение удвоенного пересечения областей к их сумме.

$$dice = \frac{2 \times TP}{(TP + FP) + (TP + FN)} \quad (4)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (5)$$

3.2 Аугментация данных во время обучения

3.2.1 Пространственные аугментации

Кроме изменения значений пикселей очевидным вариантом может стать манипуляция размерностями изображения, т.к. можно свободно менять форму и размер участков, что ещё больше разнообразит выборку. Это можно сделать следующим образом:

- Приведение к оттенкам серого - способствует неизменности цвета, модель начинает больше внимания уделять формам. Однако, в решаемой задаче множество каналов требуется для определения засеянной культуры, тогда как формы полей сравнительно одинаковые, а тип засеянной культуры не зависит от формы поля. Такая аугментация не подходит;
- Соляризация - добавляет сильный шум к цвету, чтобы модель была более устойчивой к изменению интенсивности цвета и, таким образом, больше фокусировалась на форме. Опять же, форма полей в поставленной задаче сильно совпадает, когда как волновые данные (т.н. цвет) говорят о произрастании разных культур. Не подходит;
- Размытие по Гауссу - немного меняет детали изображения. Слабо влияет на значения пикселей относительно всего изображения, но всё же помогает разнообразить выборку. Данная аугментация подходит под задачу, можно применять её с некоторой вероятностью при обучении.

Помимо этого уже использовались: отражение по горизонтали и вертикали, поворот на 5-10 градусов относительно центра изображения. Они будут использоваться и в этот раз. Вероятность применения пространственных аугментаций считается отдельно от остальных, т.е. несколько пространственных аугментаций одновременно могут быть использованы, например, с размытием.

3.2.2 Сглаживание целевых значений

Используя сглаживание целевых векторов (label smoothing), которые представлены в форме one-hot encoding, можно повысить генерализацию и

скорость обучения. Так же этот метод позволяет бороться с переобучением и «самоуверенностью» модели: например, когда модель имеет показатель точности 0.6, но при этом на предсказании выдаёт одному из классов вероятность 0.9.

Сглаживание происходит следующим образом:

- Правильным ответом на одном из вхождений выборки является класс 2, а всего классов 3. Тогда имеем следующий целевой вектор в формате one-hot encoding:

$$\vec{y}_{hot} = (0, 1, 0)$$

- Производим сглаживание по следующей формуле:

$$\vec{y}_{ls} = (1 - \alpha) * \vec{y}_{hot} + \alpha/N$$

где: α - гиперпараметр, определяющий степень сглаживания, N - количество классов. При $\alpha = 0$ сглаживания не происходит, а при $\alpha = 1$ мы получим равномерное распределение вероятностей между всеми классами.

- Приняв значение $\alpha = 0.1$, получаем вектор:

$$\vec{y}_{ls} = (0.0(33), 0.9(33), 0.0(33))$$

В данной работе был подобран коэффициент $\alpha = 0.13$ эмпирическим путём. Стоит отметить, что данный подход применим только при использовании кросс-энтропии в качестве функции ошибки при обучении.

3.2.3 Стохастическая глубина

С помощью подхода Stochastic Depth можно отсрочить переобучение и получить возможность улучшить качество модели, нарастив количество слоёв. Вторая особенность не имеет особой важности в данной задаче, так как с текущими ресурсами обучение трансформеров и так проходит на пределе.

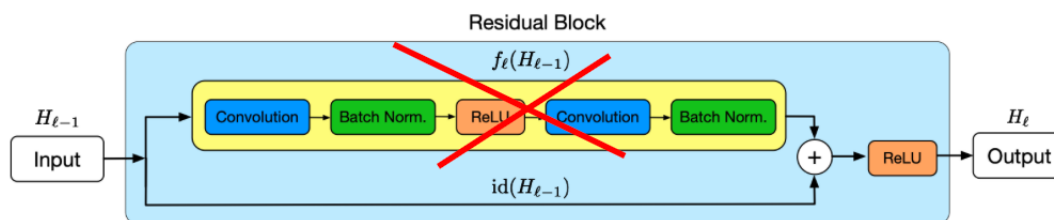


Рис. 16: Иллюстрация обнуления ветки в residual-блоке

Суть алгоритма проиллюстрирована на рис. 16: В блоках модели с обходными связями при обучении с некоторой вероятностью используются только значения с обходных веток, а значения остальных веток обнуляются. Это работает как со свёрточными блоками (например, residual), так и с блоками трансформеров.

Данный алгоритм можно выразить следующим образом:

- Есть выражение, с помощью которого вычисляется значение блока с обходными соединениями:

$$f(x) = ReLU(g(x) + x)$$

где x - входное значение, $f(x)$ - выходное, $g(x)$ - выход ветки, в которой происходит обработка (именно эта ветка и будет периодически обнуляться).

- То же выражение, но с применением предложенного подхода:

$$f(x) = ReLU(g(x) * b_l + x)$$

где b_l - случайное значение из распределения Бернулли с вероятностью $P(b_l = 1) = q$, при этом $(1 - q)$ - это вероятность, с которой мы хотим обнулять ветку.

На рис. 17 можно увидеть иллюстрацию того, как при применении данного алгоритма при росте количества слоёв уменьшается ошибка на тестовой выборке, а без применения, наоборот, растёт [16].

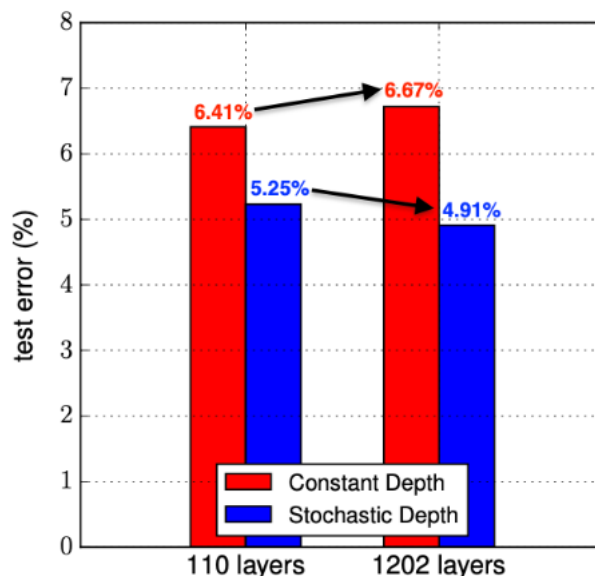


Рис. 17: Поведение ошибки при увеличении размера модели

Архитектура модели	precision	recall	F1
Модифицированный UNet	0.94	0.91	0.92
UNet3+	0.91	0.87	0.88
SwinUNet	0.92	0.88	0.90
TransUNet	0.94	0.93	0.93

Таблица 3: Результаты обучения представленных моделей на агрикультурном наборе

3.3 Анализ экспериментов

Код экспериментов был написан на языке Python с использованием Keras API с TensorFlow в качестве бэкенда. Эксперименты проводились внутри Jupyter Notebook на платформе Google Colaboratory. Обучение во всех случаях производилось с помощью алгоритма оптимизации Adam, а в качестве функции ошибки была Sparse Categorical Crossentropy. Также во время обучения отслеживались метрики Sparse Categorical Accuracy и F1. Стоит отметить, что обучение сетей с трансформерами происходило дольше классических моделей, но данные об этом не фиксировались. Различие во времени обучения было минимум в 1.5 раза в пользу сетей без трансформеров.

Усредненные значения полученных метрик указаны в таблице 3. Подробные метрики по классам указаны в приложениях 1А и 1Б.

Примеры сегментированных участков указаны на рис. 18.

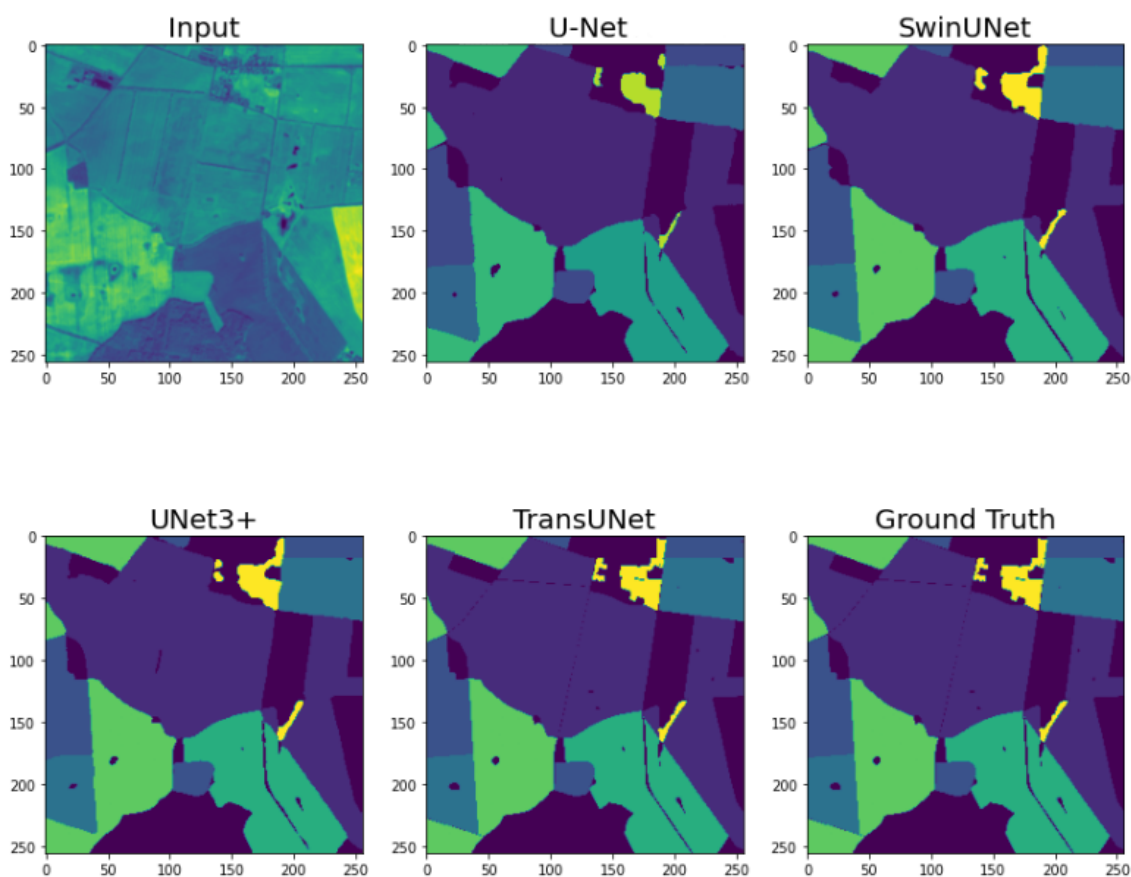


Рис. 18: Примеры сегментации каждой из моделей

По результатам наилучшей архитектурой оказалась TransUNet. Она и будет использоваться в программе.

3.4 Индекс вегетации

Normalized difference vegetation index (NDVI) — нормализованный вегетационный индекс (6), простой, но эффективный индекс для количественной оценки зеленой растительности. Он нормализует рассеивание сигнала от зеленых листьев в ближнем инфракрасном диапазоне с поглощением хлорофилла в красных длинах волн.

$$NDVI = \frac{Band_8 - Band_4}{Band_8 + Band_4} \quad (6)$$

Enhanced Vegetation Index (EVI) — расширенный вегетационный индекс (7), имеет более высокую чувствительность к биомассе, атмосферному фону и состоянию почвы. Он рассматривается как модифицированная вер-

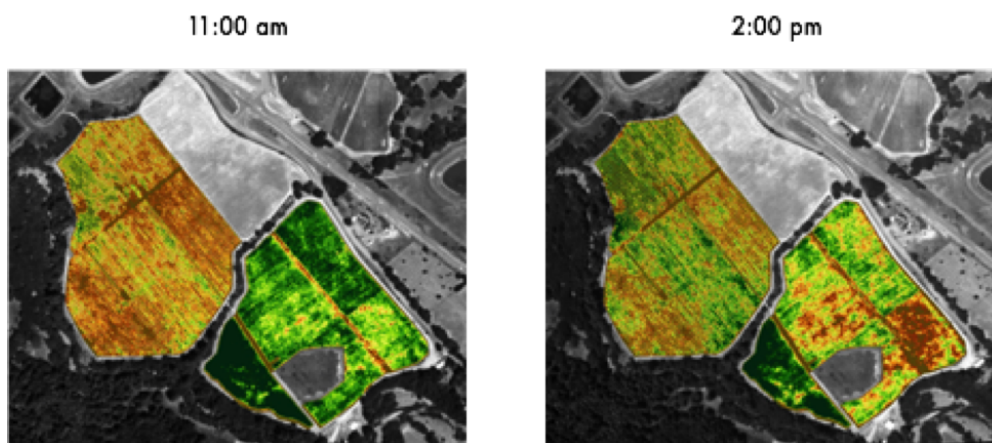


Рис. 19: NDVI, посчитанный с промежутком в 3 часа

сия NDVI с большими возможностями мониторинга растительности путем коррекции всех внешних шумов. EVI рассчитывается аналогично NDVI, но использует дополнительные длины волн сигналов для корректировки неточностей NDVI. EVI не так значительно изменяется при изменении угла падения солнца, искажениях в отраженном свете, вызванных частицами в воздухе, сигналов от наземного покрова под растительностью. Поэтому при просмотрах индексов снимков с периодом в нескольких часов можно увидеть, что EVI достаточно стабилен, когда как NDVI - нет, что может вызвать нежелательные сомнения у пользователей потенциальной системы мониторинга.

$$EVI = 2.5 \times \frac{Band_8 - Band_4}{Band_8 + 6 \times Band_4 - 7.5 \times Band_2 + 1} \quad (7)$$

Диапазон значений представленных индексов составляет от -1 до 1. Значения, приближающиеся к -1, соответствуют воде. Значения, близкие к нулю (от -0,1 до 0,1), обычно соответствуют бесплодным участкам скал, песка или снега. Невысокие положительные значения представляют кустарники и пастбища (примерно от 0,2 до 0,4), а высокие значения (приближающиеся к 1) указывают на влажные леса умеренного и тропического пояса.

В разрабатываемой программе для обработки снимков в качестве метода подсчёта индекса вегетации был взят именно EVI.

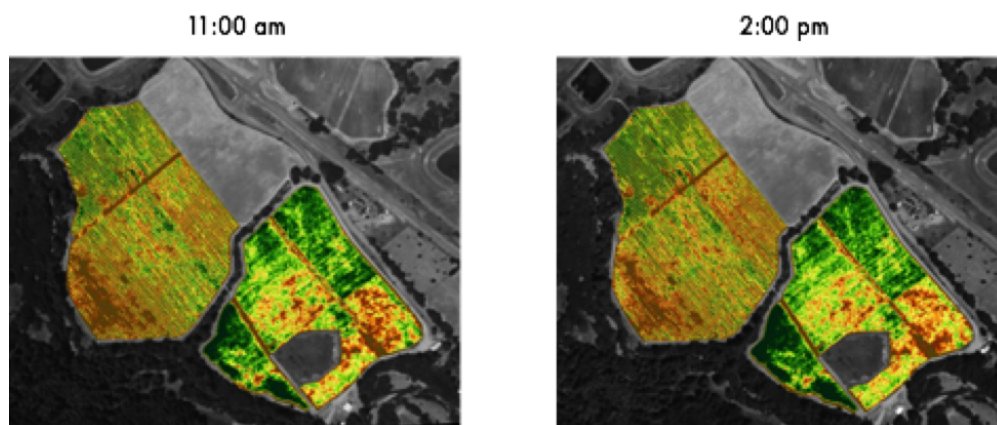


Рис. 20: EVI, посчитанный с промежутком в 3 часа

3.5 Итоговая программа

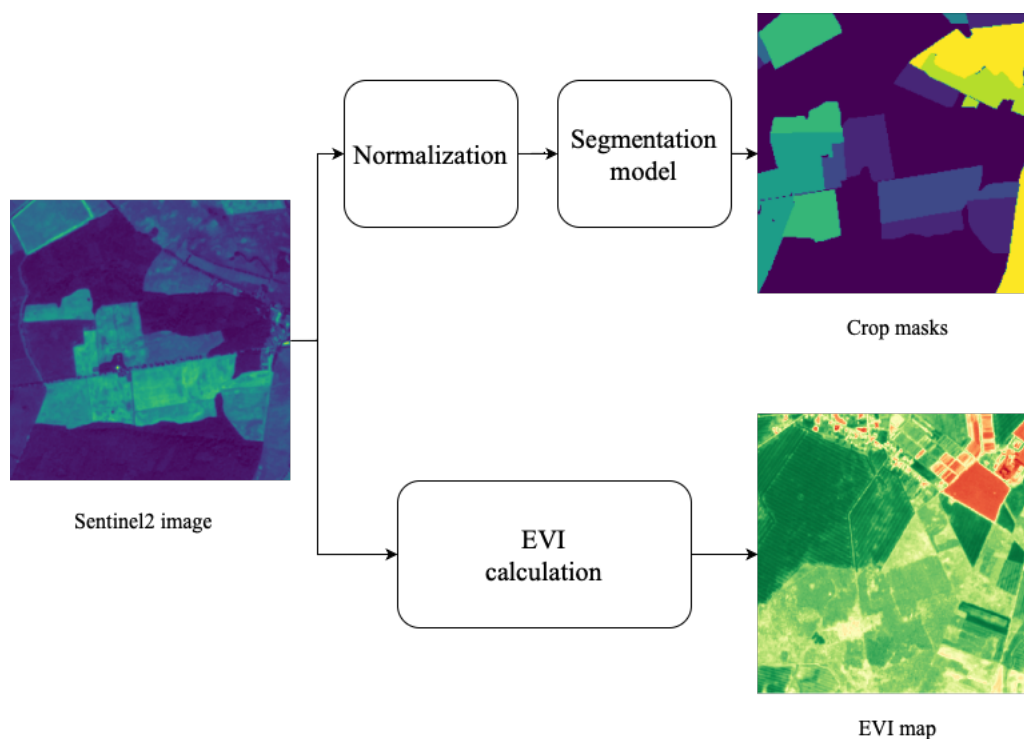


Рис. 21: Схема работы программы

Программа принимает на вход оптический снимок со спутника Sentinel2 – такой же, как в текущей тренировочной выборке. Значения снимка масштабируются в диапазон от -1 до 1 по формуле:

$$f(X) = (X - \bar{X})/sd,$$

где X - снимок, \bar{X} - среднее значение на всём датасете, sd - стандартное

отклонение на всём датасете. После этого, проходя через нейронную сеть, получаются наборы вероятностей сельскохозяйственных культур в каждом пикселе. С помощью функции $\text{argmax}()$ эти наборы преобразуются в номера классов и получается маска, которую можно увидеть в верхнем правом углу рис. 21.

Далее программа, используя тот же самый входной снимок, по формуле (7) вычисляет расширенный вегетационный индекс. Результат такого вычисления можно увидеть в нижнем правом углу рис. 21.

Заключение

В результате работы был собран и предобработан набор оптических спутниковых снимков. Было получены результаты об обучении простой модели на нескольких наборах частот (в том числе и полном). Рассмотрено несколько архитектур моделей для сегментации, проведено их обучение на собранной выборке. К тому же было проведено сравнение двух индексов вегетации, из которых выбран наиболее подходящий. Метод сегментации и вычисление индекса вегетации были обёрнуты в программу, на вход которой подаётся оптический спутниковый снимок.

Список литературы

- [1] Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. // Medical Image Computing and Computer-Assisted Intervention. MICCAI. 2015. P. 234—241
- [2] Deep learning for satellite imagery via image segmentation. [Электронный ресурс] URL: <https://deepsense.ai/deep-learning-for-satellite-imagery-via-image-segmentation/>
- [3] Landsat 8 Bands and Band Combinations [Электронный ресурс] URL: <https://gisgeography.com/landsat-8-bands-combinations/> (дата обращения: 15.04.22).
- [4] Hu Cao, Yueyue Wang et al. «Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation», 2021 (дата обращения: 01.05.2023).
- [5] Ze Liu, Yutong Lin et al. «Swin Transformer: Hierarchical Vision Transformer using Shifted Windows»
- [6] European Space Agency «MultiSpectral Instrument Overview» [Электронный ресурс] URL: <https://web.archive.org/web/20201017053209/https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument> (дата обращения: 01.05.2023).
- [7] Copernicus Open Access Hub [Электронный ресурс] URL: <https://scihub.copernicus.eu/dhus/> (дата обращения: 01.05.2023).
- [8] A Fusion Dataset for Crop Type Classification in Germany [Электронный ресурс] URL: https://mlhub.earth/data/dlr_fusion_competition_germany (дата обращения: 01.05.2023).
- [9] Н. Butler and M. Daly et al. «The GeoJSON Format», RFC 7946, 2016 (дата обращения: 01.05.2023).
- [10] Sentinel-2 mission guide [Электронный ресурс] URL: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2> (дата обращения: 15.12.22).

- [11] Alexey Dosovitskiy, Lucas Beyer et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // ArXiv, 2010.11929. 2020 (дата обращения: 01.05.2023).
- [12] Huimin Huang, Lanfen Lin et al. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation // ArXiv, 2004.08790. 2020 (дата обращения: 01.05.2023).
- [13] Jieneng Chen, Yongyi Lu et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation // ArXiv, abs/2102.04306. 2021 (дата обращения: 01.05.2023).
- [14] Н. Touvron, М. Cord and Н. Jégou «DeiT III: Revenge of the ViT», 2022 (дата обращения: 01.05.2023).
- [15] R. Müller, S. Kornblith and G. Hinton «When Does Label Smoothing Help?», 2019 (дата обращения: 01.05.2023).
- [16] G. Huang, Y. Sun et al. «Deep Networks with Stochastic Depth», 2016 (дата обращения: 01.05.2023).
- [17] Earth Observatory «Measuring Vegetation (NDVI & EVI)» [Электронный ресурс] URL: https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_1.php
- [18] Keras API [Электронный ресурс] URL: <https://keras.io>
- [19] Tensorflow, Machine Learning Framework [Электронный ресурс] URL: <https://www.tensorflow.org>
- [20] Google Colaboratory Platform [Электронный ресурс] URL: <https://colab.research.google.com>
- [21] Git-репозиторий с материалами исследований представленной работы [Электронный ресурс] URL: <https://github.com/Minyewoo/satellite-crop-type-seg-master-diploma>

Приложение 1А. Подробные метрики моделей SwinUNet и TransUNet

Метрики SwinUNet:

	precision	recall	f1-score	support
Nothing	0.94	0.96	0.95	11701033
Wheat	0.90	0.92	0.91	2509637
Rye	0.91	0.90	0.90	2014446
Barley	0.92	0.89	0.90	1306197
Oats	0.94	0.85	0.89	173872
Corn	0.92	0.88	0.90	2134317
Oil Seeds	0.92	0.91	0.91	2039977
Root Crops	0.93	0.80	0.86	34805
Meadows	0.88	0.81	0.85	1066996
Forage Crops	0.92	0.85	0.88	873824
accuracy			0.92	23855104
macro avg	0.92	0.88	0.90	23855104
weighted avg	0.92	0.92	0.92	23855104

Метрики TransUNet:

	precision	recall	f1-score	support
Nothing	0.97	0.97	0.97	11291917
Wheat	0.96	0.96	0.96	2819029
Rye	0.94	0.95	0.95	1817050
Barley	0.94	0.95	0.95	1104689
Oats	0.92	0.93	0.93	214496
Corn	0.96	0.95	0.96	2335727
Oil Seeds	0.96	0.97	0.96	2377128
Root Crops	0.92	0.85	0.88	93565
Meadows	0.86	0.88	0.87	1040061
Forage Crops	0.92	0.92	0.92	990818
accuracy			0.96	24084480
macro avg	0.94	0.93	0.93	24084480
weighted avg	0.96	0.96	0.96	24084480

Приложение 1Б. Подробные метрики модели UNet3+

	precision	recall	f1-score	support
Nothing	0.93	0.95	0.94	11938201
Wheat	0.91	0.94	0.92	3007987
Rye	0.91	0.91	0.91	2326043
Barley	0.87	0.95	0.91	1261819
Oats	0.95	0.87	0.91	325077
Corn	0.95	0.89	0.92	2719770
Oil Seeds	0.92	0.94	0.93	2349623
Root Crops	0.96	0.79	0.87	96710
Meadows	0.89	0.60	0.72	1121203
Forage Crops	0.77	0.87	0.82	1067967
accuracy			0.92	26214400
macro avg	0.91	0.87	0.88	26214400
weighted avg	0.92	0.92	0.91	26214400