Семантическая сегментация гиперспектральных изображений с использованием сверточных нейронных сетей и механизма внимания

Д.Н. Грибанов [1], А.В. Мухин [1], И.А. Килбас [1], Р.А. Паринге р [1]
[1] Самарский национальный исследовательский университет имени академика С.П. Королёва, 443086, Россия, г. Самара, Московское шоссе, д. 34

Аннотация

В работе исследуется влияние механизма внимания на точность сегментации гиперспектральных изображений сверточными нейронными сетями в области агрокультуры. Проведено исследование, в котором сравниваются две вариации нейросетевых архитектур: с использованием механизма внимания и без. Механизм внимания был реализован в виде двух модулей: позиционный и канальный. Позиционный модуль учитывает глобальный контекст, используя информацию о пространственной области всего изображения. Канальный модуль, в свою очередь, учитывает информацию всех спектральных компонент. Для проведения сравнительного исследования использовались архитектуры L2Net и U-Net. Были разработаны модифицированные версии с добавлением механизма внимания: L2AT-Net и ULAT-Net. Результаты экспериментов показали, что добавление механизма внимания в архитектуры U-Net и L2Net позволило повысить среднее значение метрики F1 с 0,80 до 0,83 и с 0,74 до 0,78 соответственно. Результаты исследования показывают, что применение механизма внимания позволяет повысить качество семантической сегментации гиперспектральных изображений.

Ключевые слова: семантическая сегментация, механизм внимания, гиперспектральные данные, нейронные сети, машинное обучение.

Цитирование: **Грибанов, Д.Н.** Семантическая сегментация гиперспектральных изображений с использованием сверточных нейронных сетей и механизма внимания / Д.Н. Грибанов, А.В. Мухин, И.А. Килбас, Р.А. Паринге р // Компьютерная оптика. – 2024. – Т. 48, № 6. – С. 894-902. – DOI: 10.18287/2412-6179-CO-1371.

Citation: Gribanov DN, Mukhin AV, Kilbas IA, Paringer RA. Semantic segmentation of hyperspectral images using convolutional neural networks and the attention mechanism. Computer Optics 2024; 48(6): 894-902. DOI: 10.18287/2412-6179-CO-1371.

Введение

Анализ гиперспектральных изображений является одной из ключевых областей дистанционного зондирования. Каждый пиксель гиперспектрального изображения представляет собой многомерный вектор, компоненты которого соответствуют различным длинам волн света в диапазоне от видимого до ближнего инфракрасного излучения. Большое количество спектральных компонент ведет к высокой информативности гиперспектральных изображений, что делает эти изображения ценным инструментом во многих областях благодаря их способности к точному различению спектральных характеристик различных объектов интереса [1 – 4].

Одной из наиболее важных задач при анализе гиперспектральных изображений является классификация [5]. В рамках задачи классификации необходимо у входного пикселя определить метку класса. Пиксель, в свою очередь, является вектором спектрального компонента. Если же классификация производится для каждого пикселя входного изображения, тогда речь идет о задаче семантической сегментации.

Классификация, соответственно и сегментация, гиперспектральных данных позволяет решить задачу определения качества еды [6], обнаружения опухоли при помощи гиперспектральных изображений языка [7], оценить урожайность пшеницы [8], а также измерить содержание хлорофилла в водоемах [9]. Однако при анализе гиперспектральных данных возникает ряд проблем: большая вариативность спектральной сигнатуры объектов интереса и малый объём обучающих данных относительно высокой размерности гиперспектральных данных. Первая проблема обусловлена вариативностью освещения, разницей в погодных условиях и т.д. [10]. Вторая проблема обусловлена дороговизной оборудования, что используется для сбора гиперспектральных данных. Так, например, известный набор снимков Земли с воздуха содержит всего 4 гиперспектральных изображения [11], в то время как для цветных изображений доступен миллионный архив ImageNet [12].

Ранние работы фокусируются на исследовании роли гиперспектральной составляющей в задаче классификации. В этих работах использовались методы попиксельной классификации с использованием полносвязанных нейронных сетей [13], а также с использованием алгоритма ближайшего соседа [1] или случайного леса [3]. В дополнение к озвученным методам также исследовались подходы к извлечению

информативных признаков и уменьшению размерности, такие как метод главных компонент (PCA) [14], метод независимых компонент (ICA) [15] и линейный дискриминантный анализ (LDA) [16]. Однако полученные классификационные маски получаются неточными, что вызвано вариативностью спектральных сигнатур объектов интереса.

На текущий момент наиболее распространенным и эффективным подходом в обработке гиперспектральных изображений являются сверточные нейронные сети (CNN) [4, 17 – 19]. Основное преимущество этих моделей заключается в их способности учитывать пространственный контекст за счет иерархической структуры, что позволяет извлекать и обрабатывать информацию на различных уровнях абстракции. Это делает CNN особенно подходящими для задач, где важно улавливать и анализировать сложные пространственные отношения, что критически важно в семантической сегментации гиперспектральных изображений. Тем не менее, применение сверточных нейронных сетей для анализа гиперспектральных изображений сталкивается с серьезными проблемами, главная из которых – ограниченное количество доступных обучающих данных. Высокая размерность гиперспектральных данных в сочетании с относительно малым объемом выборки затрудняет эффективное обучение стандартных архитектур CNN. Это создает потребность в разработке оптимизированных версий этих нейронных сетей, которые могли бы более эффективно обрабатывать гиперспектральные данные, снижая требования к объему обучающих данных и улучшая качество семантической сегментации.

Однако оптимизация архитектур CNN для анализа гиперспектральных изображений приводит к значительному ограничению их способности к учету пространственного контекста. Это связано с необходимостью уменьшения пространственных размеров ядер свертки и сокращения количества слоев в сети, чтобы адаптировать модели под ограниченные объемы данных. Данная проблема исследуется в работе [19]. Подобное ограничение становится критически важным в задачах, где необходимо улавливать сложные пространственные отношения между дальними элементами изображения, что является ключевым в семантической сегментации.

В связи с этим одним из перспективных направлений в области анализа изображений является применение механизма внимания. Механизм внимания позволяет модели фокусироваться на наиболее информативных частях изображения, учитывая при этом глобальный пространственный контекст всего изображения. Это достигается за счет динамического перераспределения весов, что позволяет сети адаптивно реагировать на различные области изображения. Механизм внимания широко применяется как в анализе цветных изображений [20 – 22], так и в анализе гиперспектральных и мультиспектральных изображений [23 – 27] в области зондирования Земли. Чаще всего механизм внимания в этих работах представляет из себя комбинацию нескольких вариаций алгоритмов вычисления внимания. Например, в статье [27] авторы используют два механизма внимания: «Модуль пространственного внимания», что выделяет важные регионы на изображении; «Модуль спектрального внимания», что оценивает важность отдельных карт признаков, при этом учитывается весь пространственный контекст карт.

В рамках данной работы проведено исследование влияния механизма внимания на качество семантической сегментации гиперспектральных изображений в области агрокультуры. Это позволило оценить потенциал применения механизма внимания в задачах обработки данных сложных и многомерных гиперспектральных изображений.

1. Позиционный и канальный механизм внимания

В данной работе предлагается использование механизма внимания для агрегации контекстной информации, позволяющего учесть глобальный контекст и улучшить точность сегментации нейросетями.

В работах [28, 29] используются два модуля механизма внимания: позиционный (Position attention module, кратко PAM) и канальный (Channel attention module, кратко CAM). Модуль PAM учитывает глобальный контекст, используя информацию о пространственной области всего изображения. Модуль CAM, в свою очередь, учитывает информацию всех спектральных компонент. Исследования показали, что их совместное использование повышает точность в задачах на основе цветных изображений. При этом данная комбинация механизма внимания была исследована в различных приложениях, от сегментации улиц [28] до медицинской диагностики [29].

Схематичное представление работы данных модулей изображено на рис. 1.

Рассмотрим принцип работы модуля PAM. Пусть имеется некоторое входное изображение $I \in \mathbb{R}^{B\times C1\times H1\times W1}$, где B – размер входной партии, C_I – количество каналов, H_I – высота, а W_I – ширина. На основе данного входного тензора нейронной сетью формируется тензор признаков $F \in \mathbb{R}^{B\times C\times H\times W}$, где C – количество каналов или карт признаков, H и W – высота и ширина соответственно. Далее для удобства будем рассматривать модуль без размерности входной партии B.

Тензор признаков F пропускается через сверточные слои для того, чтобы получить следующие тензоры $K^p \in \mathbb{R}^{(C/R)\times(H\times W)}$, $Q^p \in \mathbb{R}^{(C/R)\times(H\times W)}$ и $V^p \in \mathbb{R}^{C\times(H\times W)}$, где $R \in N$ – это скалярное ненулевое положительное значение для уменьшения количества входных признаков.

Данные карты признаков описываются как ключи (Keys), запросы (Queries) и значения (Values) соответственно. При этом пространственная размерность преобразуется в одно измерение, чтобы последующее

матричное умножение было возможно. В работах [28, 29] значение R выбирается равным 8, если значение карты признаков больше 32, иначе выбираются числа степени двойки. Основная необходимость использования подобного параметра заключается в экономии памяти при расчете матрицы внимания.

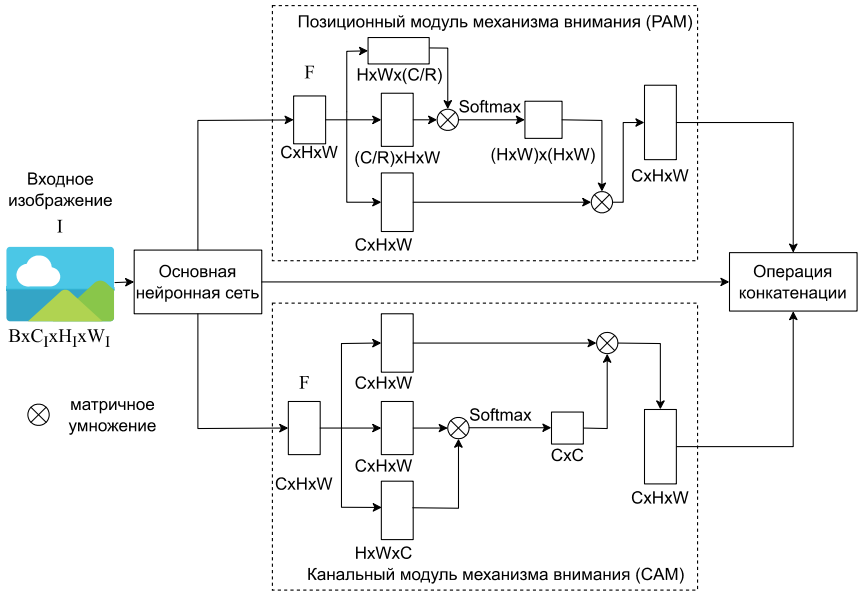

Рис. 1. Диаграмма, отображающая принципы работы позиционного и канального модуля механизма внимания

Матрица внимания $S^p \in \mathbb{R}^{(H\times W)\times(H\times W)}$ формируется матричным умножением между тензорами K^{pT} и Q^p и применением Softmax на результате. Итоговая формула имеет следующий вид:

$$S^p = Softmax\left(K^{pT} * Q^p\right). \quad (1)$$

В тензоре S^p для каждого пикселя имеется бинарная матрица внимания размерностью $H\times W$. Из-за большой размерности K^p и Q^p расчет S^p – одна из самых вычислительно сложных операций в PAM-модуле. Параметр R позволяет кратно уменьшить количество каналов C в тензорах признаков K^p и Q^p, что ведет к кратному уменьшению вычислительной сложности операции.

Итоговой тензор признаков $F^p \in \mathbb{R}^{C\times H\times W}$ формируется матричным умножением между V^p и S^p. Конечная формула имеет вид:

$$F^p = \alpha^p * \left(V^p * S^p\right), \quad (2)$$

где α^p – скалярный параметр, инициализированный нулевым значением и изменяющийся в процессе обучения.

Рассмотрим принцип работы CAM-модуля. В тензоре F по аналогии с PAM пространственная размерность преобразуется в одно измерение. Обозначим подобный тензор как $A^c \in \mathbb{R}^{C\times(H\times W)}$.

Матрица внимания $S^C \in \mathbb{R}^{C\times C}$ формируется матричным умножением между A^c и тензором A^{cT} и применением Softmax на результате. Конечная формула имеет вид:

$$S^c = Softmax\left(A^c * A^{cT}\right). \quad (3)$$

Каждая i-я строка в данной матрице представляет собой вектор вероятностей, указывающих вероятность схожести рассматриваемого признака i относительно других.

Итоговой тензор признаков $F^c \in \mathbb{R}^{C\times H\times W}$ формируется матричным умножением между F и S^c. Конечная формула имеет вид:

$$F^c = \alpha^c * \left(S^c * F\right), \quad (4)$$

где α^c – скалярный параметр, аналогичный α^p.

Результирующий тензор $F^r \in \mathbb{R}^{3C\times H\times W}$ формируется из F^p, F^c и F применением операции конкатенации. Результирующий тензор признаков F^r формируется как $F^r = contact^C(F^p, F^c, F)$, где $concat$ – это функция конкатенации, которая объединяет тензоры вдоль указанной оси C. Эта операция позволяет сохранить исходную информацию каждого тензора, что может способствовать повышению точности в задачах глубокого обучения. Хотя в работах [23, 24] и используется суммирование, однако конкатенация не приводит к потере информации за счет наложения признаков. Сравнение двух данных операций изучается в работе DenseNet [30], и, согласно ее результатам, основное преимущество в пользу конкатенации.

2. Разработка архитектуры сверточной нейронной сети

В работе [19] было показано, что в сегментации гиперспектральных изображений архитектура U-Net [31] уступает по точности модифицированной L2Net согласно метрике F1. Однако U-Net показывает худшие результаты для отдельных классов, достигающие значения нуля, что может быть связано с её склонно-

стью к переобучению из-за большого числа параметров (31 миллион против 500 тысяч у L2Net). Одним из решений в работе [19] является сокращение числа параметров.

Разработана новая архитектура ULAT-Net (U Like Attention Net), основанная на U-Net, специализированной на сегментации, но с меньшим числом параметров, равное одному миллиону, для снижения риска переобучения, сохраняя высокую точность. Уменьшение количества параметров достигается уменьшением количества сверточных слоев в блоках, а также уменьшением количества признаков в данных блоках. ULAT-Net интегрирует модули внимания CAM и PAM, сохраняет U-образную структуру U-Net и обеспечивает компактность, схожую с L2Net. Схему архитектуры можно увидеть на рис. 2. Использование модулей механизма внимания позволяет улучшить качество обработки данных за счет более точного фокусирования на важных признаках входного сигнала.

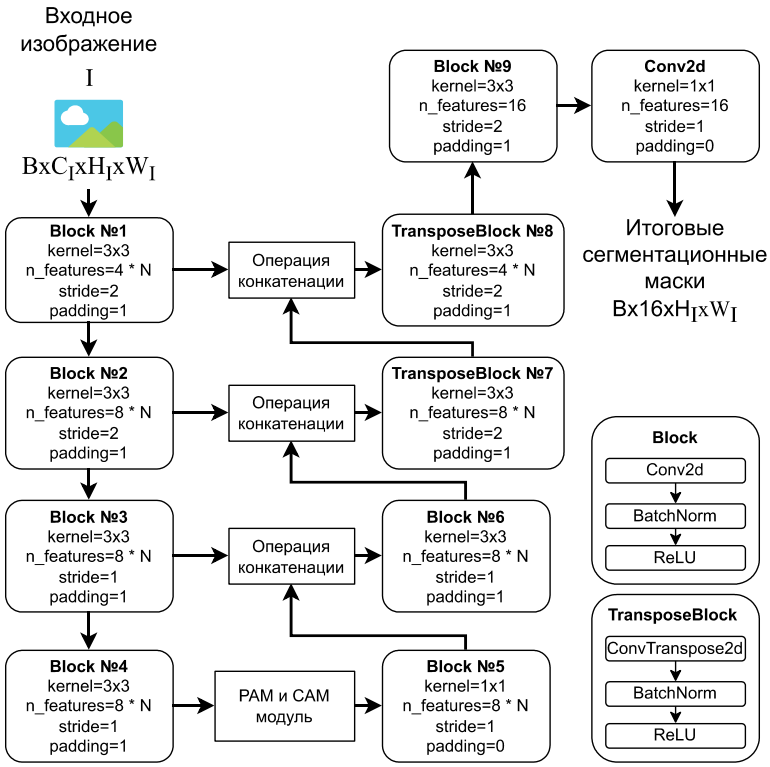

Рис. 2. Схематичное представление ULAT-Net

Разработана также архитектура UWAT-Net (U Without Attention Net). В отличие от ULAT-Net, архитектура UWAT-Net не содержит модулей CAM и PAM. Для обеспечения справедливого сравнения между двумя архитектурами и оценки чистого эффекта от применения механизмов внимания в UWAT-Net было произведено увеличение количества обучаемых параметров в блоке «block №4». Такое увеличение параметров в трехкратном размере компенсирует отсутствие CAM и PAM, позволяя всей сети поддерживать аналогичное количество обучаемых параметров, как в ULAT-Net.

Таким образом, несмотря на разницу в структуре – наличие механизмов внимания в ULAT-Net и их отсутствие в UWAT-Net, общее число обучаемых параметров в обеих архитектурах остается эквивалентным. Это обеспечивает равные условия для сравнения их работы и позволяет непосредственно оценить вклад механизмов внимания в улучшение производительности нейронной сети, исключая влияние различий в количестве обучаемых параметров.

Также в модифицированную L2Net из исследования [19] были интегрированы модули внимания для анализа их эффективности. Будут сравниваться две версии архитектуры: с механизмами внимания L2AT-Net (L2 Attention Network) и без них L2WAT-Net (L2 Without Attention Network). Модули CAM и PAM были добавлены перед последним блоком этих сетей.

3. Набор данных

В эксперименте использовался набор размеченных гиперспектральных изображений, представляющих из себя отсканированные посевы [32]. Данный набор данных состоит из 385 гиперспектральных изображений с 236 спектральными каналами, отображающими длину волны от 420 нм до 979 нм и с исходным размером 512 пикселей в ширину и высоту. Каждое гиперспектральное изображение имеет сегментированную маску. На одной маске может быть до 16 классов, не включая фоновый класс.

В наборе данных есть следующие классы: яблоня (I), свекла (II), капуста (III), морковь (IV), кукуруза

(V), огурец (VI), баклажан (VII), трава (VIII), молочай (IX), овес (X), перец (XI), картофель (XII), амарант (XIII), клубника (XIV), соя (XV) и помидор (XVI). Набор данных вручную был разделен на обучающую и тестовую выборку.

В рассматриваемом наборе данных имеется проблема дисбаланса данных. Так, например, присутствуют 67 гиперспектральных изображений с классом помидор (XVI), но с классом баклажан (VII) изображений лишь 5. Из-за этого необходимы соответствующие процедуры предобработки.

В статье [19], использующей данный набор данных, можно увидеть, что класс сои (XV) всегда имеет нулевое значение метрики F1. Это может быть связано с некорректной разметкой или высокой корреляцией класса с другими. Для проверки гипотезы была обучена нейронная сеть из упомянутой работы и проанализирована итоговая матрица ошибок.

Как и в исследовании [19], у класса сои (XV) нулевая точность, а ошибка первого рода при сравнении его с классом огурца (VI) составила 76,47 %. Это подтверждает высокую корреляцию между классами сои (XV) и огурца (VI), в связи с чем в последующих экспериментах они будут объединены в один класс. В результате общее количество классов для классификации в экспериментах сократится до 15, не учитывая фон. В связи с изменениями в наборе данных нейронная сеть из статьи [19] будет повторно обучена для последующего анализа результатов.

4. Постановка эксперимента

Предобработка гиперспектральных данных

Согласно исследованию [19], рекомендуемые размеры гиперспектрального изображения – 16, 32 или 128 пикселей с 236 каналами. Большие изображения ограничивают размер обучающей партии из-за ограничений по памяти. Аналогично рассматриваемому исследованию так же будет применяться алгоритм PCA [33] с 17 главными компонентами для уменьшения размерности данных, что позволяет увеличить объём обучающих пакетов и ускоряет обучение. Применение PCA, как показано в [19], эффективнее, чем использование гиперспектральных изображений в исходной размерности или только цветных изображений. Данная операция применяется перед аугментацией.

Для анализа нейронных сетей выбран входной размер изображения, равный 128 пикселям в ширину и высоту, обеспечивающий баланс точности и скорости обучения в соответствии с [19]. Методы аугментации взяты из той же работы и включают классические подходы, такие как случайные повороты и отражения изображения.

Исходный набор данных имеет гиперспектральные изображения размером 512 пикселей в ширину и высоту. Для получения желанного размера исходные изображения делятся на 16 пересекающихся фрагментов размером 128 пикселей в ширину и высоту каждый.

Обучающая и тестовая выборки собраны вручную и сбалансированы по классам так, чтобы в каждую часть набора попали примерно равные пропорции каждого из классов. Итоговый набор данных содержит 5120 и 656 фрагментов в обучающей и тестовой выборке соответственно.

Для каждого спектрального канала вычисляется среднее значение и среднеквадратическое отклонение (СКО) отсчета (пикселя) на основе всех гиперспектральных изображений из обучающей выборки. Перед подачей гиперспектральных изображений в нейронную сеть производится нормализация посредством поканального вычитания рассчитанного ранее среднего и деления на СКО.

Параметры обучения нейронных сетей

Количество эпох обучения равняется 70. Размер пакета равняется 32. Другие параметры, такие как оптимизатор и правило изменения обучающего шага, взяты из работы [19]. Нейронная сеть и цикл обучения реализованы с использованием фреймворка PyTorch [34]. Функцией ошибки в работе [19] является фокусная функция ошибки (Focal Loss) [35] с параметром, равным 5,5. Однако предполагается, что на текущем наборе данных фокусная функция избыточна и перекрестная энтропия покажет результаты не хуже.

Рассмотрим подробно эти функции ошибок. Пусть g – распределение истинных разметок, тогда p – распределение предсказанных меток нейронной сетью, в таком случае перекрестная энтропия определяется как:

$$L_{CE}\left(g,\,p\right)=-\mathrm{E}_{g}\left[\log\left(\mathrm{p}\right)\right]. \tag{5}$$

При обучении на наборе данных с дисбалансом в классах перекрестная энтропия приводит к неудовлетворительным результатам. Для обучения на подобных данных была разработана фокусная функция ошибки [35] с параметром γ. Параметр γ позволяет регулировать значение перекрестной энтропии, уменьшать значение для уверенных предсказаний и, наоборот, увеличивать для неуверенных предсказаний. Таким образом происходит автоматическое взвешивание ошибки, что позволяет улучшить качество предсказаний на несбалансированных наборах данных [35].

Фокусная функция ошибки имеет вид:

$$L_{F}\left(g,p\right)=-\mathrm{E}_{g}\left[\left(1-p\right)^{\gamma}\cdot\log\left(\mathrm{p}\right)\right]. \tag{6}$$

При равенстве параметра γ нулю функция ошибки представляет собой перекрестную энтропию.

Для сравнения результатов эффективности двух функций ошибок будет проведено два обучения архитектуры L2WAT-Net с двумя разными функция ошибок. Лучшая функция будет использована далее при обучении архитектур на основе L2Net и U-Net с вариациями с механизмом внимания и без.

5. Обсуждение результатов

В табл. 1 отражены результаты экспериментов. В данной таблице показан средний результат по нескольким запускам обучения с использованием метрики F1 для каждого класса, среднее значение и взвешенное среднее.

Среднее значение метрики F1 рассчитывается как арифметическое среднее по всем классам, придавая каждому из них равное значение. Взвешенное среднее метрики F1 учитывает частоту встречаемости классов в данных, умножая F1 каждого класса на его долю в наборе, после чего значения суммируются, давая взвешенное значение метрики F1.

Из таблицы видно, что обучения L2WAT-Net перекрестной энтропией в качестве функции ошибки имеет среднее значение метрики F1, равное 0,74, что на 0,03 больше, чем обучение с фокусной функцией ошибки, равное 0,71. Таким образом, влияние дисбаланса классов несущественно для использования фокусной ошибки как основной функции для обучения.

Табл. 1. Результаты экспериментов с использованием метрики F1

	L2WAT-Net (Focal loss)	L2WAT-Net	L2AT-Net	UWAT-Net	ULAT-Net
Фоновый класс	0,84	0,86	0,84	**0,87**	0,86
I	0,80	0,81	0,79	**0,85**	**0,85**
II	0,95	0,95	0,95	**0,96**	**0,96**
III	0,77	0,79	0,74	**0,83**	0,80
IV	0,40	0,41	0,59	0,60	**0,69**
V	0,39	0,48	0,70	0,59	**0,72**
VI	0,59	0,61	0,75	0,74	**0,81**
VII	0,61	0,64	0,72	0,75	**0,80**
VIII	0,70	0,71	0,85	0,85	**0,92**
IX	0,59	0,65	0,70	**0,71**	**0,71**
X	0,72	0,75	0,80	0,81	**0,83**
XI	0,87	0,89	0,84	**0,92**	0,91
XII	0,87	0,88	0,88	**0,91**	**0,91**
XIII	**0,93**	0,92	0,91	0,92	**0,93**
XIV	0,80	0,82	0,78	0,84	**0,86**
XV	0,59	0,60	0,58	0,65	**0,68**
Среднее	0,71	0,74	0,78	0,80	**0,83**
Взвешенное среднее	0,81	0,82	0,84	0,86	**0,87**

При сравнении результатов обучения сетей с архитектурами L2WAT-Net и L2AT-Net видно, что добавление CAM- и PAM-модулей повышает среднее значение метрики F1 на 0,04 и на 0,02 при взвешенном среднем.

Архитектура UWAT-Net имеет значение метрики F1, равное 0,8, в то время как L2WAT-Net и L2AT-Net имеют значение, равное 0,74 и 0,78 соответственно. Добавление механизма внимания в UWAT-Net архитектуру увеличивает итоговое значение F1 до 0,83. Таким образом, разработанная архитектура ULAT-Net с механизмом внимания имеет точность лучше, чем разработанная архитектура L2WAT-Net из статьи [19].

Визуальный анализ матриц внимания, генерируемых механизмами CAM и PAM

Проанализируем влияние использования модулей CAM и PAM, а также рассмотрим результирующие карты признаков из данных модулей на основе трех гиперспекральных изображениий, которые можно увидеть на рис. 3.

Каждая строка состоит из гиперспектрального изображения в RGB-пространстве, двух карт признаков из CAM- и PAM-модуля, исходной разметки, результата сегментации посредством использования UWAT-Net и ULAT-Net. На данном рисунке #12 означает, что рассматривается 12-я карта признаков. В разметке черному соответствует регион земли, в то время как ярким цветом обозначены растения.

Рассмотрим подробно результат на третьей строке для PAM-модуля. Видно, что данный модуль выделяет достаточно точно важные отдельные регионы классов согласно истинной разметке. Подобное можно заметить как на #1, так и на #12. Так, например, в третьем столбце в случае #1 выделяется зона растения слева, а в случае #12 выделяется зона земли.

Структура PAM-модуля позволяет выделить релевантные области, усиливая вклад информативных зон и подавляя менее значимые участки. Визуализация полученных карт признаков из данного модуля демонстрирует концентрацию отдельных карт признаков на отдельных классах объектов, что свидетельствует о различимости интереса сети к разным частям изображения и классам.

Хотя интерпретация CAM-модуля представляет сложность при рассмотрении ранее формул данного модуля, поскольку в отличие от PAM-модуля CAM-модуль не создает отдельную матрицу внимания для каждого пикселя. CAM-модуль устанавливает взаимосвязи между каналами, моделируя их корреляцию и позволяя сети адаптивно уточнять сигналы на уровне каналов, что способствует улучшению дифференциации классов.

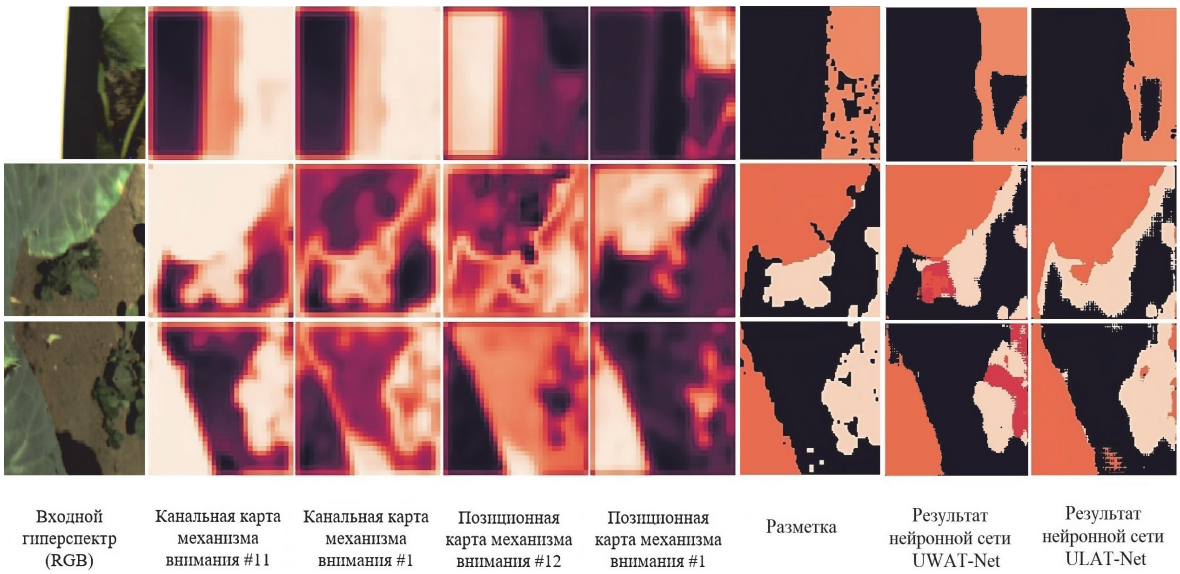

Рис. 3. Карты внимания и маски для трех изображений из набора данных

Рассмотрим результаты на третьей строке для CAM-модуля. На #1 видно, что выделяется земля (класс фон, черный регион) и зона у растения сверху, в то время как на #11 выделяются большими значениями лишь растения, тем самым эти две карты признаков выделяют отдельные группы классов.

Из рассмотренных карт признаков можно увидеть, что CAM-модуль группирует отдельные классы, тем самым фокусируя внимание нейронной сети на отдельных кластерах классов.

Таким образом PAM-модуль более точно выделяет отдельные значимые регионы, тогда как CAM-модуль ориентирован на группы классов. Такое различие в подходах модулей иллюстрирует комплементарность механизмов внимания в задаче семантической сегментации.

Применение механизма внимания в архитектуре ULAT-Net способствует более корректной сегментации классов, что подтверждается сравнением результатов с UWAT-Net на рис. 3. Так, например, в строке 2 видно, что UWAT-Net допускает ошибку и появляется лишний класс при сегментации класса растения посередине, а также при сегментации растения в строке 3 справа, в то время как ULAT-Net не имеет подобной ошибки.

Стоит отметить, что отдельные контуры растений лучше выделяются с использованием UWAT-Net, чем с использованием ULAT-Net. Так, например, в строке 2 видно, как при использовании ULAT-Net появляется излишняя разметка класса посередине на полученном результате нейронной сети. В случае UWAT-Net подобная ошибка лишь встречается сверху изображения, но снизу излишней разметки нет. Аналогичное можно заметить на рисунке как в строке 1, так и в строке 3.

Таким образом, несмотря на схожие показатели по взвешенной метрике F1 у архитектур ULAT-Net и UWAT-Net, присутствие механизма внимания в ULAT-Net позволяет получить более корректную сегментацию классов растений на входных изображениях. Рассматривая результаты с использованием архитектуры U-WAT-Net, стоит отметить, что она лучше справляется с выделением контуров объектов.

Заключение

В работе исследовано влияние механизма внимания на точность сегментации гиперспектральных изображений сверточными нейронными сетями в области агрокультуры. Механизм внимания был внедрен в известные архитектуры сверточных нейронных сетей: U-Net и L2Net. Для реализации механизма внимания использовались два модуля: позиционный (PAM) и канальный (CAM).

Были разработаны архитектуры на основе U-Net: UWAT-Net без механизма внимания и ULAT-Net с использованием механизма внимания, сочетающего в себе модули PAM и CAM. На основе L2Net была разработана архитектура L2AT-Net также с использованием механизма внимания, сочетающего в себе модули PAM и CAM.

Результаты экспериментов показали, что добавление механизма внимания в архитектуры U-Net и L2Net позволило повысить среднее значение метрики F1 с 0,80 до 0,83 и с 0,74 до 0,78 соответственно. Результаты исследования показывают, что применение механизма внимания в виде двух модулей PAM и CAM позволяет повысить качество семантической сегментации гиперспектральных изображений.

Благодарности

Результаты исследования были получены при поддержке государственного задания Минобрнауки России в рамках исследования, выполненного лабораторией Самарского университета «Фотоника для

умного дома и умного города» в рамках проекта № FSSS-2021-0016 (теоретическая часть и разработка технологии) и за счет средств государственного задания в сфере научной деятельности (проект FSSS-2024-0014) (программная реализация).

References

[1] Fabelo H, Ortega S, Ravi D, et al. Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations. PLOS ONE 2018; 13(3): e0193721. DOI: 10.1371/journal.pone.0193721.

[2] Paringer RA, Mukhin AV, Kupriyanov AV. Formation of an informative index for recognizing specified objects in hyperspectral data. Computer Optics 2021; 45(6): 873-878. DOI: 10.18287/2412-6179-CO-930.

[3] Amini S, Homayouni S, Safari A, Darvishsefat AA. Object-based classification of hyperspectral data using Random Forest algorithm. Geo-Spat Inf Sci 2018; 21: 127-138. DOI: 10.1080/10095020.2017.1399674.

[4] Paoletti ME, Haut JM, Plaza J, Plaza A. A new deep convolutional neural network for fast hyperspectral image classification. ISPRS J Photogramm Remote Sens 2018; 145: 120-147. DOI: 10.1016/j.isprsjprs.2017.11.021.

[5] Grewal R, Singh Kasana S, Kasana G. Machine learning and deep learning techniques for spectral spatial classification of hyperspectral images: A comprehensive survey. Electronics 2023; 12: 488. DOI: 10.3390/electronics12030488.

[6] Leiva-Valenzuela GA, Lu R, Aguilera JM. Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. J Food Eng 2013; 115: 91-98. DOI: 10.1016/j.jfoodeng.2012.10.001.

[7] Li Z, Wang H, Li Q. tongue tumor detection in medical hyperspectral images. Sensors 2011; 12: 162-174. DOI: 10.3390/s120100162.

[8] Liu LY, Wang JH, Huang WJ, Zhao CJ, Zhang B, Tong QX. Improving winter wheat yield prediction by novel spectral index. Trans CSAE 2004; 20: 172-175.

[9] Kutser T, Paavel B, Verpoorter C, Kauer T, Vahtmäe E. Remote sensing of water quality in optically complex lakes. Int Arch Photogramm Remote Sens Spatial Inf Sci 2012; XXXIX-B8: 165-169. DOI: 10.5194/isprsarchives-xxxix-b8-165-2012.

[10] Ishihara M, Inoue Y, Ono K, Shimizu M, Matsuura S. The impact of sunlight conditions on the consistency of vegetation indices in croplands–effective usage of vegetation indices from continuous ground-based spectral measurements. Remote Sens 2015; 7: 14079-14098. DOI: 10.3390/rs71014079.

[11] Graña M. Hyperspectral remote sensing scenes – Grupo de Inteligencia Computacional (GIC). Wwwehueus 2011. Source: <http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes>.

[12] Fei-Fei L, Deng J, Li K. ImageNet: Constructing a large-scale image database. J Vision 2010; 9: 1037. DOI: 10.1167/9.8.1037.

[13] Li J, Bioucas-Dias JM, Plaza A. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. IEEE Trans Geosci Remote Sens 2010; 48(11): 4085-4098. DOI: 10.1109/tgrs.2010.2060550.

[14] Prasad S, Bruce LM. Limitations of principal components analysis for hyperspectral target recognition. IEEE Geosci Remote Sens Lett 2008; 5: 625-629. DOI: 10.1109/lgrs.2008.2001282.

[15] Villa A, Benediktsson JA, Chanussot J, Jutten C. Hyperspectral image classification with independent component discriminant analysis. IEEE Trans Geosci Remote Sens 2011; 49: 4865-4876. DOI: 10.1109/tgrs.2011.2153861.

[16] Bandos TV, Bruzzone L, Camps-Valls G. Classification of hyperspectral images with regularized linear discriminant analysis. IEEE Tran Geosci Remote Sens 2009; 47: 862-873. DOI: 10.1109/tgrs.2008.2005729.

[17] Ran L, Zhang Y, Wei W, Zhang Q. A hyperspectral image classification framework with spatial pixel pair features. Sensors 2017; 17: 2421. DOI: 10.3390/s17102421.

[18] Trajanovski S, Shan C, Weijtmans PJC, de Koning SGB, Ruers TJM. Tongue tumor detection in hyperspectral images using deep learning semantic segmentation. IEEE Trans Biomed Eng 2021; 68(4): 1330-1340. DOI: 10.1109/tbme.2020.3026683.

[19] Mukhin A, Danil G, Paringer R. semantic segmentation of hyperspectral imaging using convolutional neural networks. Optical Memory and Neural Networks 2022; 31: 38-47. DOI: 10.3103/s1060992x22050071.

[20] Kirillov A, Mintun E, Ravi N, et al. Segment anything. arXiv Preprint. 2023. Source: <https://arxiv.org/abs/2304.02643>. DOI: 10.48550/arxiv.2304.02643.

[21] Chen B, Liu YQ, Zhang Z, Lu G, Kong AWK. TransAttUnet: Multi-level attention-guided U-Net with transformer for medical image segmentation. IEEE Trans Emerg Top Comput Intell 2023; 8(1): 55-68. DOI: 10.1109/tetci.2023.3309626.

[22] Vanian V, Zamanakos G, Pratikakis I. Improving performance of deep learning models for 3D point cloud semantic segmentation via attention mechanisms. Computers & Graphics 2022; 106: 277-287. DOI: 10.1016/j.cag.2022.06.010.

[23] Han Z, Hong D, Gao L, Yao J, Zhang B, Chanussot J. Multimodal hyperspectral unmixing: Insights from attention networks. IEEE Trans Geosci Remote Sens 2022; 60: 5524913. DOI: 10.1109/tgrs.2022.3155794.

[24] Shi C, Liao D, Zhang T, Wang L. Hyperspectral image classification based on 3D coordination attention mechanism network. Remote Sens 2022; 14: 608. DOI: 10.3390/rs14030608.

[25] Mohla S, Pande S, Banerjee B, Chaudhuri S. FusAtNet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and LiDAR classification. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2020: 416-425. DOI: 10.1109/cvprw50498.2020.00054.

[26] Zheng X, Chen W, Lu X. Spectral super-resolution of multispectral images using spatial–spectral residual attention network. IEEE Trans Geosci Remote Sens 2022; 60: 5404114. DOI: 10.1109/tgrs.2021.3104476.

[27] Zhang H, Yao J, Li N, Gao L, Huang M. Multimodal attention-aware convolutional neural networks for classification of hyperspectral and LiDAR data. IEEE J Sel Top Appl Earth Obs Remote Sens 2023; 16: 3635-3644. DOI: 10.1109/jstars.2022.3187730.

[28] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2019: 3141-3149. DOI: 10.1109/cvpr.2019.00326.

[29] Sinha A, Dolz J. Multi-scale self-guided attention for medical image segmentation. IEEE J Biomed Health Inform 2021; 25: 121-130. DOI: 10.1109/jbhi.2020.2986926.

[30] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. arXiv Preprint. 2016. Source: <https://arxiv.org/abs/1608.06993>. DOI: 10.48550/arxiv.1608.06993.

[31] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. arXiv Preprint. 2015. Source: <https://arxiv.org/abs/1505.04597>. DOI: 10.48550/arxiv.1505.04597.
[32] HSI-Dataset-API: API for accessing HSI datasets. 2021. Source: <https://pypi.org/project/HSI-Dataset-API>.
[33] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom Intell Lab Syst 1987; 2: 37-52. DOI: 10.1016/0169-7439(87)80084-9.
[34] Paszke A, Gross S, Massa F, et al. PyTorch: An imperative style, high-performance deep learning library. Proc 33rd Int Conf on Neural Information Processing Systems 2019: 8026-8037.
[35] Lin T-Y, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. IEEE Trans Pattern Anal Mach Intell 2018; 42(2): 318-327. DOI: 10.1109/tpami.2018.2858826.

Сведения об авторах

Грибанов Данил Николаевич, 2000 года рождения, студент факультета информатики Самарского национального исследовательского университета имени академика С.П. Королева (Самарский университет), старший лаборант научно-исследовательской лаборатории автоматизированных систем научных исследований. Круг научных интересов включает интеллектуальный анализ данных, распознавание образов и искусственные нейронные сети. E-mail: *gribanov.dn@ssau.ru*

Мухин Артем Владимирович, 1999 года рождения, студент факультета информатики Самарского национального исследовательского университета имени академика С.П. Королева (Самарский университет), старший лаборант научно-исследовательской лаборатории автоматизированных систем научных исследований. Круг научных интересов включает интеллектуальный анализ данных, распознавание образов и искусственные нейронные сети. E-mail: *mukhin.av@ssau.ru*

Килбас Игорь Александрович, 2000 года рождения, студент факультета информатики Самарского национального исследовательского университета имени академика С.П. Королева (Самарский университет), старший лаборант научно-исследовательской лаборатории автоматизированных систем научных исследований. Круг научных интересов включает интеллектуальный анализ данных, распознавание образов и искусственные нейронные сети. E-mail: *kilbas.ia@ssau.ru*

Парингер Рустам Александрович, 1990 года рождения, доцент кафедры технической кибернетики Самарского национального исследовательского университета имени академика С.П. Королева (Самарский университет). В 2013 году окончил факультет информатики СГАУ. Кандидат технических наук с 2017 года. Круг научных интересов включает интеллектуальный анализ данных, распознавание образов и искусственные нейронные сети. E-mail: *rusparinger@ssau.ru*

ГРНТИ: 28.23.15
Поступила в редакцию 14 июня 2023 г. Окончательный вариант – 8 апреля 2024 г.

Semantic segmentation of hyperspectral images using convolutional neural networks and the attention mechanism

D.N. Gribanov [1], A.V. Mukhin [1], I.A. Kilbas [1], R.A. Paringer [1]
[1] Samara National Research University,
443086, Samara, Russia, Moskovskoye Shosse 34

Abstract

This paper investigates an effect of the attention mechanism on the accuracy of hyperspectral image segmentation by convolutional neural networks in agriculture. The study compares two modifications of neural network architectures: with and without the attention mechanism. The attention mechanism is implemented as two modules: position-based (PAM) and channel-based (CAM). The positional module (PAM) considers the global context using information about the spatial domain of the whole image. The channel module (CAM) in turn takes into account the information of all spectral components. L2Net and U-Net architectures are used for a comparative study. Modified versions with the addition of the attention mechanism are developed: L2AT-Net and ULAT-Net. The experimental results show that adding the attention mechanism to the U-Net and L2Net architectures increases the mean value of the F1 metric from 0.80 to 0.83 and from 0.74 to 0.78, respectively. The results show that the application of the attention mechanism can improve the quality of semantic segmentation of hyperspectral images.

Keywords: semantic segmentation, attention mechanism, hyperspectral data, neural network, machine learning.

Citation: Gribanov DN, Mukhin AV, Kilbas IA, Paringer RA. Semantic segmentation of hyperspectral images using convolutional neural networks and the attention mechanism. Computer Optics 2024; 48(6): 894-902. DOI: 10.18287/2412-6179-CO-1371.

Acknowledgements: The work was partly funded by the Russian Federation Ministry of Science and Higher Education under the state project FSSS-2021-0016, “Photonics for a smart home and smart city” (theoretical part and technology development) and under the state research project FSSS-2024-0014) (software implementation).

Authors’ information

Danil Nikolaevich Gribanov, (b. 2000), master's student Informatics faculty at Samara National Research University, senior lab assistant of the research laboratory “Photonics for a Smart Home and Smart City”. Research interests include data mining, computer vision and artificial neural networks. E-mail: *gribanov.dn@ssau.ru*.

Artem Vladimirovich Mukhin, (b. 1999), master's student Informatics faculty at Samara National Research University, senior lab assistant of the research laboratory “Photonics for a Smart Home and Smart City”. Research interests include computer vision, artificial neural networks, real-time high-load systems. E-mail: *mukhin.av@ssau.ru*.

Igor Alexandrovich Kilbas, (b. 2000), master's student Informatics faculty at Samara National Research University, senior lab assistant of the research laboratory “Photonics for a Smart Home and Smart City”. Research interests include data mining, artificial neural networks and language models. E-mail: *kilbas.ia@ssau.ru*.

Rustam Alexandrovich Paringer, (born 1990), received Master’s degree in Applied Mathematics and Informatics from Samara State Aerospace University (2013). He received his PhD in 2017. Associate professor of the Technical Cybernetics department of Samara National Research University. Research interests: data mining, machine learning and artificial neural networks. E-mail: *rusparinger@ssau.ru*.

Received June 14, 2023. The final version – April 8, 2024.