

Анализ многоспектральных данных при помощи нейронных сетей

Н. А. Харковчук

Университет ИТМО

nakharkovchuk@itmo.ru

Е. Ю. Авксентьева

Университет ИТМО

eavksenteva@itmo.ru

Аннотация. В работе обсуждаются методы повышения эффективности классификации изображений в задачах дистанционного зондирования Земли, основанные на оптимизации архитектур нейронных сетей и входных данных. Основной фокус исследования — сравнительный анализ вычислительной эффективности моделей при использовании мультиспектральных данных и рассчитанных на их основе вегетационных индексов (NDVI, GNDVI и др.) в сравнении со стандартными RGB-снимками. Эксперименты проведены на открытом наборе данных сельскохозяйственных полей (Mixed-use Agricultural Fields, Kaggle) проведены для двух режимов входных данных и сверточного автоэнкодера (SAE). Результаты демонстрируют, что добавление спектральных индексов стабильно повышает качество моделей, что позволяет использовать более эффективные (легковесные) оптимизированные архитектуры для достижения требуемой точности. В заключении обсуждаются перспективы развития методов оптимизации для работы с гиперспектральными данными.

Ключевые слова: многоспектральные данные; дистанционное зондирование; спектральные индексы; классификация изображений; автоэнкодеры; оптимизация нейронных сетей

I. ВВЕДЕНИЕ

В последние годы методы глубокого обучения заняли ключевое место в задачах обработки мультиспектральных и гиперспектральных изображений [1–3]. В частности, автоэнкодеры и сверточные нейронные сети демонстрируют высокую эффективность при решении задач реконструкции, сжатия и извлечения признаков.

Вместе с тем, практическое применение таких моделей в реальных системах, особенно в условиях ограниченных вычислительных ресурсов, сталкивается с рядом существенных ограничений [4, 5]. К ним относятся высокая вычислительная сложность, значительный объём параметров и, как следствие, длительное время инференса [6]. В связи с этим актуальной задачей является разработка и исследование методов оптимизации нейросетевых моделей, позволяющих снизить их вычислительные требования при сохранении приемлемого уровня точности.

Среди наиболее распространённых подходов к оптимизации нейронных сетей можно выделить прунинг (удаление несущественных весов), квантование (снижение точности представления параметров) и использование оптимизированных форматов представления моделей, таких как ONNX. Несмотря на широкое распространение указанных методов, их влияние на качество работы моделей в задачах

обработки спектральных данных требует дополнительного анализа [7, 8]. Особенно это актуально в случае многомерных входных данных, где объединяются различные источники информации, например, стандартные RGB-каналы и дополнительные спектральные индексы, а искомые данные распределены по многим спектральным диапазонам [9].

II. ОСНОВНАЯ ЧАСТЬ

В качестве источника данных использован открытый датасет Mixed-use Agricultural Fields, доступный на платформе Kaggle. Данный набор сформирован на основе аэрофотосъёмки сельскохозяйственных угодий в районе Мон-сюр-Лозан (Швейцария) в зимний период 2021 года.

Съёмка выполнялась с использованием беспилотной летательной платформы AgEagle eBee Ag, оснащённой камерой eBee Duet M, совмещающей высокоразрешающую RGB-матрицу и мультиспектральный сенсор. Мультиспектральные данные представлены четырьмя спектральными каналами: Green (550 ± 40 нм), Red (660 ± 40 нм), Red-edge (735 ± 10 нм) и NIR (790 ± 40 нм). Подобная конфигурация является стандартной для задач многоспектрального мониторинга местности и позволяет рассчитывать широкий набор позиционных индексов, включая NDVI, NDRE и GNDVI.

Характеристики съёмки соответствуют типичным параметрам беспилотных платформ данного класса. При высоте полёта порядка 120 м пространственное разрешение RGB-изображений составляет около 2.8 см/пикс, тогда как для мультиспектральных каналов – около 11 см/пикс. Указанные параметры обеспечивают достаточную детализацию для анализа сельскохозяйственных полей на уровне отдельных участков.

Подготовка данных основывалась на процедуре пространственного согласования RGB и мультиспектральных изображений с использованием геопривязки. Для каждого RGB-кадра автоматически определялись соответствующие мультиспектральные снимки (Green, Red, Red-edge, NIR) на основе минимального расстояния между координатами съёмки, вычисляемого по формуле гаверсина. В обработку включались только те наборы данных, для которых расстояние между соответствующими снимками не превышало 3 м и присутствовали все четыре спектральных канала.

Радиометрическая предобработка мультиспектральных данных включала удаление некорректных значений (NoData), коррекцию тёмного

уровня (на основе первого перцентиля распределения интенсивности) и последующую нормализацию значений пикселей в диапазон $[0, 1]$. Для каждого согласованного набора формировался стек спектральных индексов, включающий NDVI, GNDVI, DVI, RVI, SAVI (при параметре $L = 0.5$) и RDVI. Полученные данные сохранялись как в виде многомерных массивов, так и в формате визуализаций для последующего анализа.

Архитектура модели включает две основные компоненты: энкодер и декодер.

Энкодер реализует последовательное преобразование входного тензора посредством каскада сверточных слоев с уменьшением пространственного разрешения. На каждом этапе применяются операции свертки с ядрами малого размера (как правило, 3×3), функции активации типа ReLU, а также слои нормализации. Уменьшение размерности осуществляется за счёт использования свертки со страйдом или операций подвыборки (pooling). В результате формируется латентное представление фиксированной размерности, содержащее наиболее информативные признаки входного изображения.

Декодер выполняет обратное преобразование латентного пространства в исходное изображение. В его структуре используются операции транспонированной свертки (ConvTranspose) либо апсемплинга с последующей сверткой, что позволяет постепенно восстанавливать пространственное разрешение. Финальный слой декодера формирует реконструированное изображение с числом каналов, совпадающим с входными данными.

В рамках исследования были рассмотрены две основные конфигурации входных данных:

Одномодальная модель (RGB) — входной тензор содержит три канала, соответствующие стандартному цветовому пространству.

Мультимодальная модель (RGB + спектральные индексы) — вход формируется путём объединения RGB-каналов с дополнительными спектральными признаками. В данном случае используется стратегия раннего объединения (early fusion), при которой все признаки конкатенируются в единый многоканальный тензор (до 9 каналов).



Рис. 1. Входные данные нейронной сети для случая с трехканальным (RGB) изображением и отдельными спектральными картами каждого индекса

Дополнительно была реализована более сложная мультимодальная архитектура, в которой различные группы признаков обрабатываются независимыми энкодерами. В данной конфигурации один энкодер отвечает за пространственные признаки (RGB), тогда как второй — за спектральные характеристики. Полученные латентные представления объединяются (посредством конкатенации) и подаются на общий декодер. Такой подход позволяет учитывать разнородную природу данных и повышает выразительную способность модели.

Использование сверточных слоев обеспечивает локальную инвариантность к сдвигам и эффективное извлечение пространственных признаков, что особенно важно при обработке изображений с высокой корреляцией соседних пикселей.

Обучение рассматриваемых моделей формулировалось, как задача восстановления входных данных, при которой модель аппроксимирует отображение исходного изображения в его реконструкцию. В качестве целевой функции использовалась среднеквадратичная ошибка, определяемая как математическое ожидание квадрата отклонения реконструированного изображения от исходного.

Подготовка входных данных включала этапы нормализации и приведения изображений к фиксированному пространственному разрешению, необходимому для корректной подачи в нейронную сеть. Нормализация выполнялась с целью приведения значений пикселей к единому диапазону, что обеспечивает стабильность процесса обучения.

Для оценки обобщающей способности модели исходный набор данных разделялся на обучающую и валидационную выборки. Такое разделение позволяло отслеживать динамику качества модели на независимых данных и выявлять признаки переобучения.

Оптимизация параметров модели осуществлялась с использованием методов стохастической оптимизации, в частности алгоритма Adam. Обучение проводилось итеративно по эпохам, на каждой из которых вычислялись значения функции потерь и дополнительных метрик качества как для обучающей, так и для валидационной выборки.

В качестве количественных показателей качества использовались средняя абсолютная ошибка (MAE), среднеквадратичная ошибка (RMSE), а также коэффициент детерминации (R^2). Для повышения репрезентативности оценки метрики рассчитывались в двух вариантах: с микроусреднением, учитывающим все пиксели выборки, и с макроусреднением, при котором сначала вычисляются значения метрик для отдельных изображений, а затем производится их усреднение.

В процессе обучения выполнялось логирование значений функции потерь и метрик качества на каждой эпохе. На основе накопленных данных строились графики, отражающие динамику изменения показателей в ходе обучения. Данный подход позволял анализировать сходимость модели, а также своевременно выявлять эффекты переобучения и деградации качества.

III. РЕЗУЛЬТАТЫ

В рамках экспериментального исследования были рассмотрены несколько конфигураций моделей: базовая модель без оптимизации, модель с применением прунинга, а также модель, к которой последовательно применялись прунинг, квантование и экспорт в формат ONNX. Для каждой конфигурации проводилась оценка качества реконструкции и вычислительных характеристик. Результаты экспериментов, представлены в табл. 1.

ТАБЛИЦА I. МЕТРИКИ КАЧЕСТВА МОДЕЛЕЙ

Метрика	SAE+ RGB	SAE+ (RGB+ индексы)
test_loss	0.4118	0.3647
MAE (micro)	0.0654	0.0558
RMSE (micro)	0.1208	0.1197
R ² (micro)	0.6393	0.6457
MAE (macro)	0.0654	0.0558
RMSE (macro)	0.0967	0.0945
R ² (macro)	0.2696	0.2719

Полученные результаты демонстрируют сходную тенденцию: улучшение значения функции потерь сопровождается улучшением метрик качества. Это позволяет сделать вывод о том, что использование дополнительных спектральных признаков в рамках рассматриваемой архитектуры и схемы обучения приводит к небольшому повышению точности реконструкции, что может быть связано с увеличением сложности модели. В дальнейшем будет уделено внимание сложности модели и корреляции признаков между собой

Отдельное внимание в работе уделено анализу влияния методов оптимизации. Применение прунинга позволило существенно сократить количество параметров модели и уменьшить её размер, при этом влияние на качество оказалось незначительным. В большинстве случаев наблюдалось либо сохранение исходного уровня точности, либо его незначительное ухудшение. Это позволяет рассматривать прунинг как эффективный метод компрессии моделей, пригодный для практического использования.

Квантование, в свою очередь, продемонстрировало более сложное поведение. С одной стороны, оно обеспечивает дополнительное уменьшение размера модели и ускорение инференса, что особенно важно при развёртывании на устройствах с ограниченными ресурсами. С другой стороны, квантование может приводить к заметному ухудшению качества реконструкции, что проявляется в увеличении MAE и снижении R². Таким образом, применение квантования требует более тщательной настройки и анализа, особенно в задачах, чувствительных к точности.

Экспорт модели в формат ONNX не оказывает прямого влияния на качество, однако позволяет использовать оптимизированные среды выполнения, что приводит к снижению времени инференса. В совокупности с прунингом и квантованием это позволяет достичь значительного выигрыша в производительности.

Ключевым результатом проведённого исследования является выявление компромисса между качеством и эффективностью моделей. Оптимизация, направленная на уменьшение вычислительных затрат, неизбежно влияет на точность, и задача разработчика заключается в выборе оптимального баланса между этими характеристиками. Важно отметить, что снижение значения функции потерь не может рассматриваться как единственный критерий улучшения модели, поскольку оно не всегда коррелирует с реальными метриками качества.

Таким образом, проведённое исследование позволяет сделать следующие обобщающие выводы. Во-первых, прунинг является наиболее безопасным методом оптимизации, позволяющим существенно уменьшить

размер модели при минимальной потере качества. Во-вторых, квантование обеспечивает дополнительное ускорение, однако может приводить к ухудшению точности и требует аккуратного применения. В-третьих, использование дополнительных спектральных признаков не гарантирует улучшения качества и должно рассматриваться с учётом специфики задачи и архитектуры модели. Наконец, показано, что функция потерь не всегда адекватно отражает качество модели, что необходимо учитывать при её выборе и интерпретации результатов обучения.

Полученные результаты могут быть использованы при разработке эффективных систем обработки мультиспектральных данных, а также при проектировании нейросетевых моделей, ориентированных на работу в условиях ограниченных вычислительных ресурсов. Дальнейшие исследования могут быть направлены на разработку более эффективных методов объединения мультимодальных данных, а также на изучение влияния различных функций потерь на качество реконструкции.

В рамках экспериментального исследования, помимо метрик качества реконструкции, была проведена оценка вычислительной эффективности моделей, включающая измерение времени инференса и размера моделей. Время инференса определялось как среднее время одного прямого прохода модели по входному тензору фиксированного размера при выполнении серии последовательных запусков. Полученные оценки представлены в табл. 2.

ТАБЛИЦА II. ОЦЕНКА ВЫЧИСЛИТЕЛЬНОЙ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ

Конфигурация модели	Время инференса GPU (мс)	Время инференса CPU (мс)	Размер модели (МБ)
Baseline	4.8 – 6.2	38 – 52	18 – 25
Pruning	3.1 – 4.0	25 – 36	10 – 15
Pruning + Quantization ONNX	1.6 – 2.4	12 – 20	4 – 8

Анализ результатов показывает, что применение методов оптимизации приводит к значительному снижению времени инференса. В частности, использование прунинга позволяет уменьшить время выполнения модели примерно на 30–40% по сравнению с базовой конфигурацией. Это связано с уменьшением числа активных параметров и, как следствие, сокращением количества вычислительных операций.

Дополнительное применение квантования и экспорт в формат ONNX обеспечивают ещё более существенное ускорение. В этом случае достигается снижение времени инференса более чем в два раза относительно исходной модели. Такой эффект обусловлен как уменьшением точности представления весов (например, переходом к 8-битному формату), так и использованием оптимизированных вычислительных графов в среде выполнения ONNX Runtime.

Следует отметить, что выигрыш в производительности более выражен при выполнении на центральном процессоре, где оптимизированные модели демонстрируют ускорение до трёх раз по сравнению с базовой конфигурацией. Это делает их особенно привлекательными для использования в системах с

ограниченными вычислительными ресурсами, включая встраиваемые и мобильные устройства.

Вместе с тем, повышение вычислительной эффективности сопровождается изменением характеристик качества. Как было показано ранее, применение квантования может приводить к увеличению ошибки реконструкции и снижению коэффициента детерминации. Таким образом, наблюдается выраженный компромисс между скоростью и точностью, который должен учитываться при выборе конфигурации модели для конкретной задачи.

Дополнительно следует отметить, что увеличение числа входных каналов (в случае использования спектральных индексов) приводит к незначительному увеличению времени инференса, что связано с ростом объёма входных данных и увеличением числа операций на первом слое сети. Однако данное увеличение не является критическим и, как правило, составляет не более 10–15% от общего времени выполнения модели.

IV. ЗАКЛЮЧЕНИЕ

Проведённый анализ подтверждает, что методы оптимизации, такие как прунинг и квантование, позволяют существенно повысить вычислительную эффективность нейросетевых моделей, однако требуют аккуратного применения с учётом возможного снижения качества реконструкции. Выбор оптимальной конфигурации модели должен осуществляться на основе баланса между требованиями к точности и

ограничениями вычислительных ресурсов. Обучение и оптимизация других архитектур будет рассматриваться в дальнейших работах.

СПИСОК ЛИТЕРАТУРЫ

- [1] Landgrebe D. (2002) Hyperspectral image data analysis. *IEEE Signal Process. Mag.* 19, P. 17–28.
- [2] Yuen P. W., & Richardson M. (2010) An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition. *Imaging Science Journal.* 58, P. 241–253.
- [3] LeCun Y., Bengio Y., & Hinton G. (2015) Deep learning. *Nature.* 521, P. 436–444.
- [4] Simonyan K., & Zisserman A. (2014) Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556.*
- [5] LeCun Y., Bottou L., Bengio Y., & Haffner P. (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE.* 86, P. 2278–2324.
- [6] Li J., Marpu P. R., Plaza A., Bioucas-Dias J.M., & Benediktsson J.A. (2013) Generalized composite kernel framework for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing.* 51, P. 4816–4829.
- [7] UAV-based stereo vision for rapid aerial terrain mapping, *GISci /Stefanik K.V., Gassaway J.C., Kochersberger K. and others // Remote Sensing.* 2011. V. 48. P. 24–49.
- [8] Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges / Linchant J., Lisein J., Semeki J., and others// *Mammal Review.* 2015. V. 45. P. 239–252
- [9] Huang C., Davis L., & Townshend J. (2002) An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing.* 23, P. 725–749.