

Д. А. Шатилов¹, С. А. Рылов¹✉

Оценка эффективности методов аугментации для обучения CNN-моделей на малых мультиспектральных выборках

¹Федеральный исследовательский центр информационных и вычислительных технологий,
г. Новосибирск, Российская Федерация
e-mail: RylovS@mail.ru

Аннотация. Применение моделей глубокого обучения к данным дистанционного зондирования Земли (ДЗЗ) существенно осложняется проблемой адаптации домена (domain adaptation), обусловленной статистическими различиями в обучающей и целевой выборках. В данной работе исследуется специфика и эффективность методов аугментации мультиспектральных данных для обучения свёрточных нейронных сетей (CNN) на примере задачи построения карт сегментации древостоев по данным со спутников «WorldView-2» и «Ресурс-П». Использование процедуры кросс-валидации позволило количественно оценить эффективность динамической аугментации: зафиксировано улучшение метрик на 6,17% и 14,46% для macro-accuracy и macro-IoU, соответственно. Установлено, что наибольшую эффективность демонстрируют геометрические преобразования, в то время как модификации спектральных характеристик могут приводить к деградации качества моделей. Эксперимент по трансферному обучению продемонстрировал, что спектральные трансформации повышают адаптивность моделей к межсенсорным различиям.

Ключевые слова: адаптация домена, аугментация данных, семантическая сегментация, мультиспектральные изображения, дистанционное зондирование, свёрточные нейронные сети

D. A. Shatilov¹, S. A. Rylov¹✉

Evaluating augmentation method effectiveness for CNN training on small multispectral datasets

¹Federal Research Center for Information and Computational Technologies,
Novosibirsk, Russian Federation
e-mail: RylovS@mail.ru

Abstract. The application of deep learning models to remote sensing data is significantly complicated by the domain adaptation problem, caused by statistical differences between training and target samples. This paper investigates the specificity and effectiveness of augmentation methods for multispectral data in training convolutional neural networks (CNN) on the task of tree species segmentation with «WorldView-2» and «Resurs-P» satellite data. The use of cross-validation procedure allowed quantitative evaluation of dynamic augmentation effectiveness: macro-accuracy and macro-IoU improved by 6,17% and 14,46% respectively. Geometric transformations were proved to demonstrate the highest effectiveness, while modifications of spectral characteristics can lead to model performance degradation. The transfer learning experiment demonstrated that spectral transformations increase model adaptability to inter-sensor differences.

Keywords: domain adaptation, data augmentation, semantic segmentation, multispectral images, remote sensing, convolutional neural networks

Введение

Проблема адаптации домена представляет собой одну из наиболее актуальных задач машинного обучения [1], возникающую при применении моделей, обученных на исходном наборе данных, к целевому домену с отличным статистическим распределением признаков. Современные подходы к решению данной проблемы классифицируются на три категории [2]: нормализация входных данных, методы самообучения для извлечения доменно-инвариантных признаков и использование масштабных наборов данных. Благодаря этим подходам современные архитектуры, такие как DINOv2, MAE, SAM и SegFormer, способны формировать робастные признаки, обобщаемые на различные домены.

При решении задач семантической сегментации изображений ДЗЗ выявляется ряд специфических ограничений: критический дефицит размеченных данных, разнородность спутниковых сенсоров с различным количеством спектральных каналов, а также вариабельность физической природы пиксельных данных в зависимости от уровня предобработки [3] – это могут быть, как и стандартные цифровые значения (Digital Number, DN), так и значения коэффициента отражения (Remote Sensing Reflectance, RSR).

К числу наиболее сложных относятся задачи распознавания объектов со сходными спектральными характеристиками, например, различных видов деревьев [4]. Дифференциация по типам древесных насаждений возможна преимущественно в мультиспектральном диапазоне [5], что делает невозможным прямое использование предобученных RGB-моделей.

В данной работе проводится оценка эффективности методов аугментации данных для повышения точности работы нейронных сетей при обучении на малых мультиспектральных выборках. Исследуются три категории аугментаций: пространственные, спектральные и смешивающие преобразования. Целью исследования является определение оптимальных стратегий аугментации для задач мультиспектральной сегментации в условиях ограниченного объема размеченных данных ДЗЗ.

Методы и материалы

В качестве тестового полигона выбрана территория Бронницкого лесничества в Московской области, включающая 11 доминантных классов лиственных и хвойных пород, а также классы «Не лес» и «Нет информации» (исключён из прогнозирования). Векторная карта разметки была получена в работе [6].

В настоящей работе использованы разновременные мультиспектральные спутниковые снимки для выбранной территории. Данные со спутника «WorldView-2» (съёмка выполнена в летний период до 2017 года) включают 8 спектральных каналов в диапазоне 400–1040 нм, подвергнуты процедуре атмосферной коррекции и представлены значениями RSR в формате float. Данные со спутника «Ресурс-П» (съёмка выполнена в 2018 году) включают 4 спектральных канала (RGB+NIR), подвергнуты процедуре радиометрической коррекции и представлены стандартными цифровыми значениями (Digital Numbers, DN) в формате uint16. После процедуры географи-

ческого выравнивания сформированы два датасета по 81 тайл размерами 256×256 пикселей.

Для проведения семантической сегментации применены архитектуры FPN, LinkNet, U-Net, U-Net++, Attention U-Net и U²-Net с энкодером ResNet-50 [7]. Функция потерь – линейная комбинация взвешенной функции потерь Дайса и модифицированной кросс-энтропии [8]. Оценка качества осуществлялась по метрикам maseo-accuracy и maseo-IoU. Дополнительно для оценки точности моделей использовалась 4-кратная кросс-валидация. Входные данные подвергнуты поканальной стандартизации.

Всего исследованы три категории аугментации данных:

1. Пространственные преобразования (к изображениям и маскам): трансляция, поворот, отражение.

2. Спектральные преобразования (только изображения): гауссовский шум, варьирование контраста и повышение резкости. Также предложено обобщение HSV-преобразований на мультиспектральный случай: выбор трёх случайных каналов, линейное масштабирование к диапазону $[0; 1]$, случайное преобразование тона, яркости или насыщенности с последующим обратным масштабированием.

3. Батч-смешивающие преобразования (изображения и маски): CutMix [9] и MixUp [10].

Результаты

Рассмотрены два подхода к применению аугментации: статический, предполагающий однократное расширение тренировочной выборки, и динамический, заключающийся в аугментации на каждой эпохе обучения. Обучение всех CNN-архитектур проводилось на 8-канальных снимках «WorldView-2» (RSR). Сравнивались конфигурации: статическое расширение в 4 и 8 раз, пространственные аугментации, комбинации пространственных, спектральных и батч-смешивающих преобразований. Результаты усреднены по всем моделям и кросс-валидации (рис. 1). Лучшее значение maseo-IoU достигло 0,48 при использовании динамического подхода.

В контексте проблемы адаптации домена исследован перенос весов моделей между разнородными сенсорами без дообучения: модели, обученные на данных «WorldView-2», применялись к данным спутника «Ресурс-П». Для обеспечения корректности переноса отобраны 4 канала $[R, G, B, NIR1]$, близких по спектральному диапазону. Поканальная стандартизация выполнена независимо для каждого датасета. При использовании геометрических преобразований в сочетании со спектральными и батч-смешивающими преобразованиями получены результаты, приведённые на рис. 2. Наилучшее качество достигнуто при комбинации геометрических преобразований с гауссовым шумом, варьированием контраста, повышением резкости, HSV-преобразованиями и CutMix (рис. 3).

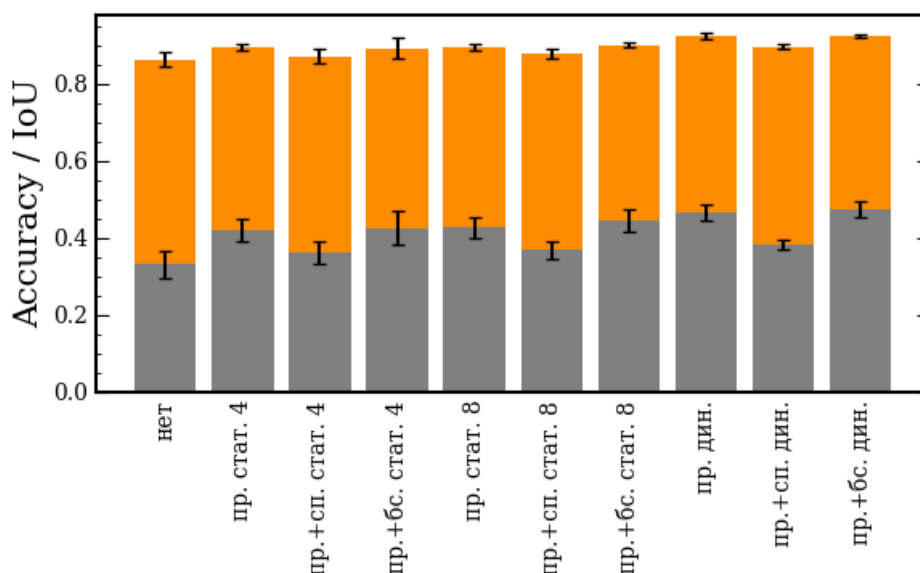


Рис. 1. Усредненные значения метрик в зависимости от различных конфигураций аугментаций при обучении на 8-канальном датасете «WorldView-2». Сокращения: пр. – пространственные, сп. – спектральные, бс. – батч-смешивающие, стат. – статические, дин. – динамические

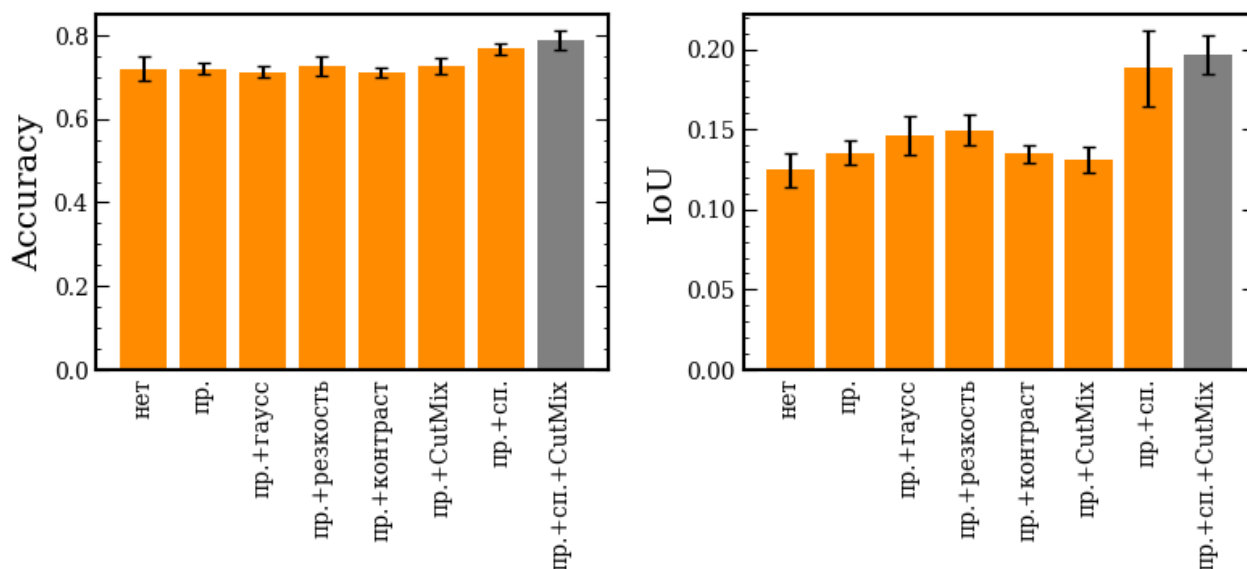


Рис. 2. Усредненные значения метрик в зависимости от различных конфигураций аугментаций, использованных во время обучения модели, при переносе ее весов на 4-канальный датасет «Ресурс-П»

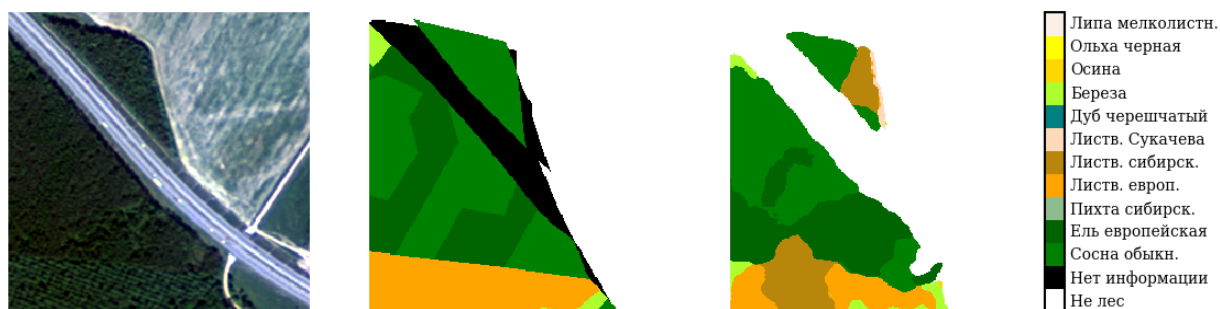


Рис. 3. RGB-композит снимка со спутника «Ресурс-П» (слева), данные разметки (по центру) и построенная Link-Net карта сегментации (справа)

Обсуждение

Значительный прирост точности (+6,17% macro-accuracy и +14,46% macro-IoU) при использовании динамической аугментации объясняется расширением разнообразия тренировочного набора и внесением стохастичности в процесс обучения, что повышает регуляризующий эффект и снижает риск переобучения [11]. Статический подход демонстрирует меньший прирост точности вследствие меньшей вариативности между эпохами, что может приводить к запоминанию артефактов аугментации и формированию ложных корреляций.

Снижение метрик при переходе к данным «Ресурс-П» обусловлено шумами синего канала и временным несоответствием между съёмкой и разметкой (вырубки, антропогенные объекты). Тем не менее, полученные результаты подтверждают корректность предложенных методов аугментации для задач семантической сегментации мультиспектральных изображений, о чём свидетельствует устойчивое уверенное распознавание классов хвойных и лиственных насаждений.

Заключение

В рамках исследования проведена оценка эффективности методов аугментации для обучения CNN на малом мультиспектральном датасете из 81 тайла размером 256×256 . Экспериментально подтверждено, что динамическая аугментация обеспечивает статистически значимое улучшение качества сегментации на 14,46% по метрике macro-IoU по сравнению с обучением без аугментации. Установлено, что пространственные преобразования вносят наибольший вклад в повышение качества моделей, тогда как спектральные преобразования могут снижать точность вследствие формирования физически некорректных комбинаций спектральных значений. При переносе весов между данными «WorldView-2» и «Ресурс-П» без дообучения показано, что комбинация независимой поканальной стандартизации и спектральных аугментаций повышает робастность моделей к межсенсорной вариации ценой некоторого снижения точности на исходных данных, что представляет собой классический компромисс между точностью и обобщающей способностью [2].

В качестве перспективного направления исследований представляется применение методов самообучения, в частности контрастивного обучения, для которых

агрессивные аугментации являются механизмом формирования позитивных пар изображений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Singhal P., Walambe R., Ramanna S., Kotecha K. Domain Adaptation: Challenges, Methods, Datasets, and Applications // IEEE Access. – 2023. – Vol. 11. – P. 6973-7020.
2. Wang M. A Deep visual domain adaptation: A survey // Neurocomputing. – 2018. – Vol. 312. – P. 135-153.
3. Hill V. J., Zimmerman R. C., Bissett P. et al. Impact of Atmospheric Correction on Classification and Quantification of Seagrass Density from WorldView-2 Imagery // Remote Sensing. – 2023. – Vol. 15, № 19. – Art. 4715. – DOI: 10.3390/rs15194715.
4. Key T., Warner T., McGraw J. et al. Comparison of Multispectral and Multitemporal Information in High Spatial Resolution Imagery for Classification of Individual Tree Species in a Temperate Hardwood Forest // Remote Sensing of Environment. – 2001. – Vol. 75 – P. 100-112.
5. Fassnacht F.E., Latifi H., Sterenczak K. et al. Review of studies on tree species classification from remotely sensed data // Remote Sensing of Environment. – 2016. – Vol. 186. – P. 64–87.
6. Дмитриев Е.В., Кондранин Т.В., Мельник П.Г., Донской С.А. Определение видового состава смешанного леса на основе совместной обработки публичных спутниковых карт и многовременных изображений Sentinel-2 // Современные проблемы дистанционного зондирования Земли из космоса. – 2023. – Т. 21. – № 1. – С. 31-50.
7. Agrawal S., Rewaskar V., Agrawal R. et al. Advancements in NSFW Content Detection: A Comprehensive Review of ResNet-50 Based Approaches // Journal of intelligent systems and applications in engineering. – 2023. – Vol. 11 (4). – P. 41-45.
8. Шатилов Д.А., Рылов С.А. Семантическая сегментация лесных массивов на спутниковых снимках высокого разрешения с помощью U-Net-подобных моделей // Вычислительные технологии. – 2025. – Т. 30, № 3. – С. 111-126.
9. Yun S., Han D., Oh S. J. et al. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features [Электронный ресурс]. – 2019. – Режим доступа: <https://arxiv.org/abs/1905.04899> (дата обращения: 02.05.2026).
10. Zhang H., Cisse M., Dauphin Y. N. et al. mixup: Beyond Empirical Risk Minimization [Электронный ресурс]. – 2018. – Режим доступа: <https://arxiv.org/abs/1710.09412> (дата обращения: 02.05.2026).
11. Shorten C., Khoshgoftaar T. M. A Survey on Image Data Augmentation for Deep Learning // Journal of Big Data. – 2019. – Vol. 6, № 1. – Art. 60.

© Д. А. Шатилов, С. А. Рылов, 2026