УДК 001.817+

СЕГМЕНТАЦИЯ СПУТНИКОВЫХ ИЗРБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ SEGFORMER

И. А. Душко[1)], А. М. Недзьведь[2), 3)], А. М. Белоцерковский[3)]

[1)] *МИРЭА – Российский Технологический Университет,*
Москва, Россия, dushko.i.a@edu.mirea.ru
[2)] *Белорусский государственный университет,*
Минск, Беларусь, nedzveda@tut.by
[3)] *Объединённый институт проблем информатики НАН Беларуси,*
Минск, Беларусь, alex.@bsu.by

В данной статье рассматривается задача сегментации спутниковых изображений с использованием модели SegFormer на основе трансформеров. Для подготовки данных использовались ручное аннотирование с использованием Labelbox и общедоступные картографические ресурсы (Google Maps, Ersi и Bing). Модели на основе трансформеров демонстрируют превосходство над традиционными свёрточными нейронными сетями (CNN) в задачах компьютерного зрения, что делает их перспективными для сегментации спутниковых изображений.

Ключевые слова: семантическая сегментация; спутниковые изображения; нейронные сети; SegFormer; трансформеры.

SEGMENTATION OF SATELLITE IMAGES USING THE MODEL SEGFORMER

I. A. Dushko[a)], A. M. Nedzved[b), c)], A. M. Belotserkovsky[c)]

[a)] *MIREA – Russian Technological University,*
Moscow, Russia, dushko.i.a@edu.mirea.ru
[b)] *Belarusian State University,*
Minsk, Belarus, nedzveda@tut.by
c) Joint Institute of Informatics Problems of the National Academy of Sciences of Belarus,
Minsk, Belarus, alex.@bsu.by

This article discusses the problem of segmentation of satellite images using the SegFormer model based on transformers. Manual annotation using Labelbox and publicly available cartographic resources (Google Maps, Ersi, and Bing) were used to prepare the data. Transformer-based models demonstrate superiority over traditional convolutional neural networks (CNNs) in computer vision tasks, which makes them promising for segmentation of satellite images.

Keywords: semantic segmentation; satellite images; neural networks; SegFormer; transformers.

1. Введение

Семантическая сегментация спутниковых изображений имеет важное значение в городском планировании, сельском хозяйстве и мониторинге окружающей среды.

Цель исследования – разработка и оценка подхода к сегментации спутниковых изображений с применением модели SegFormer, включая подготовку пользовательских наборов данных, оптимизацию процесса обучения и анализ производительности модели для достижения высокой точности сегментации.

Задачи включают:

- обзор современных архитектур сегментации;
- подготовку пользовательских наборов данных;
- анализ точности и устойчивости результатов.

2. Обзор методов

Семантическая сегментация спутниковых изображений предполагает классификацию каждого пикселя по категориям (здания, дороги, растительность). Эта техника объединяет компьютерное зрение, дистанционное зондирование и машинное обучение для извлечения информации из изображений высокого разрешения. Применения включают градостроительство (определение границ зданий), экологический мониторинг (наблюдение за вырубкой лесов) и сельское хозяйство (анализ состояния урожая). В семантической сегментации первых успехов добились Fully Convolutional Networks (FCN), заменившие полносвязные слои на сверхточные для сохранения пространственной информации, но при этом плохо локализующие мелкие объекты. Затем появились UNet с симметричной энкодер-декодер архитектурой, объединяющей высокоуровневые и низкоуровневые признаки.

Переход к трансформерам начался с Vision Transformer (ViT), доказавшего эффективность self-attention для глобального захвата взаимосвязей пикселей, а SegFormer пошёл дальше, сочетая лёгкую трансформерную структуру с мульти-масштабным извлечением признаков

Фреймворк SegFormer состоит из двух основных модулей: иерархического кодировщика Transformer для извлечения грубых и тонких признаков, и легкого декодера All-MLP для непосредственного слияния этих многоуровневых признаков и прогнозирования маски семантической сегментации (рис. 1). “FFN” обозначает сеть прямой связи.

Помимо архитектурных инноваций, ключевыми остаются задачи создания и разметки датасетов: ручное аннотирование (Labelbox, Esri, Bing

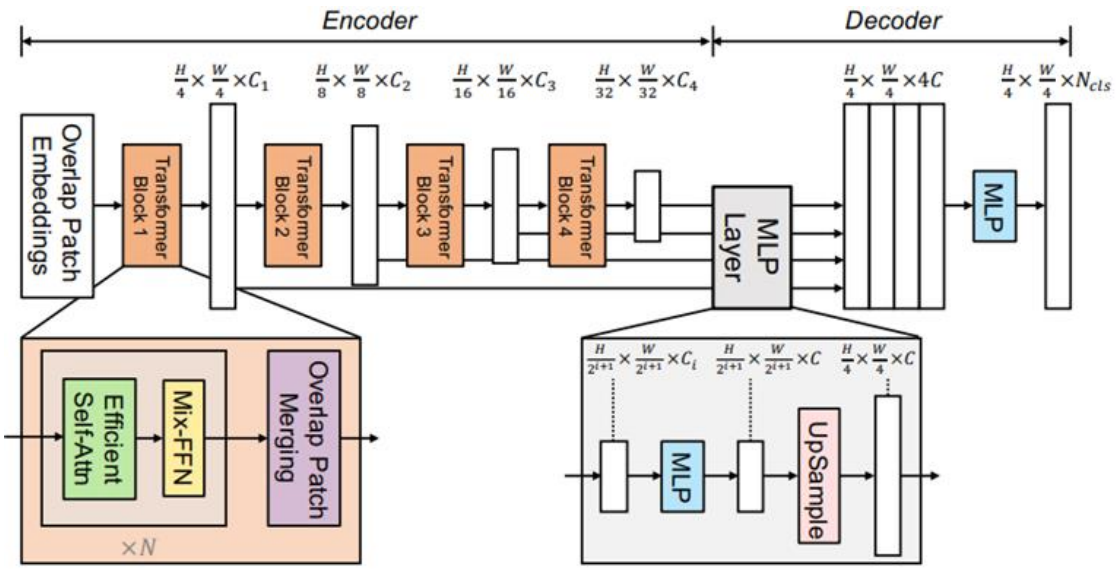

Рис. 1. Фреймворк SegFormer

Maps), балансировка классов и продвинутые аугментации (вращение, отражение, вариации цветового пространства) для увеличения разнообразия выборки и предотвращения переобучения. Практическая значимость методов подтверждена в приложениях от медицинской сегментации до анализа спутниковых снимков для мониторинга землепользования, где SegFormer показывает высокую устойчивость к шуму, мельчайшим деталям и изменению масштаба объектов.

3. Характеристика данных

Наборы данных разделены на обучающую, валидационную и тестовую выборки, как показано в табл. 1.

Таблица 1

Обзор наборов данных

Название датасета	Общее количество изображений	Обучающая выборка	Валидационная выборка	Тестовая выборка	Количество категорий
Labelbox 1	856	650	150	56	11
Labelbox 2	984	700	140	144	11
Google maps 0	956	600	178	178	36
Google maps 1	368	280	44	44	36
Google maps 2	228	180	24	24	36
Google maps 3	1284	800	242	242	36

Столбчатая диаграмма (рис. 2) иллюстрирует размеры наборов данных, включая общее количество изображений и их разбиение на обучающую, валидационную и тестовую выборки. Это позволяет оценить масштаб каждого набора данных и соотношение между выборками, что важно для понимания их пригодности для обучения моделей. Например, набор данных Google Maps 3 содержит 1284 изображения, что делает его наиболее крупным, тогда как Google Maps 2 с 228 изображениями является наименьшим.

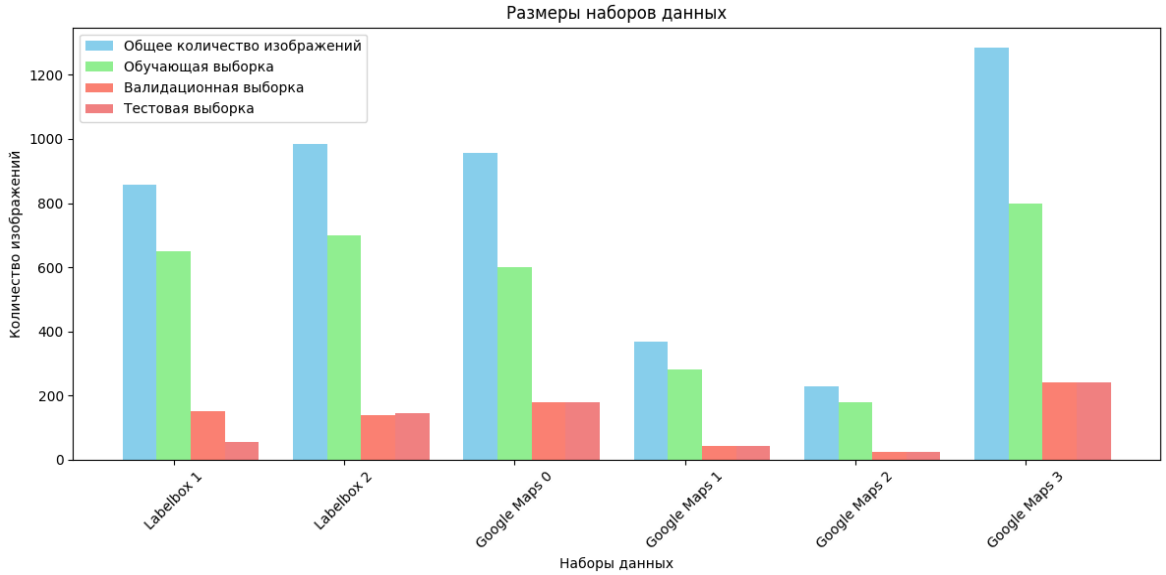

Рис. 2. Размеры наборов данных

Структура

В ходе работы было использовано два разных подхода к аннотированию наборов данных.

1. Ручное аннотирование выполнено с использованием Labelbox. Полученный набор данных состоит из пар изображений – исходных спутниковых изображений и соответствующих им масок сегментации. На рис. 3 показаны примеры вручную аннотированных пар, демонстрирующих исходные изображения вместе с соответствующими им масками сегментации.

2. Создание набора данных из общедоступных карт: во втором подходе используются изображения из общедоступных картографических ресурсов, таких как Google Maps, Esri и Bing. Эти изображения были обработаны для создания масок сегментации на основе доступных данных. Метод использует огромное количество общедоступных изображений для создания набора данных, подходящего для обучения моделей сегментации. На рис. 4 представлены примеры пар изображений, созданных с использованием этого метода.

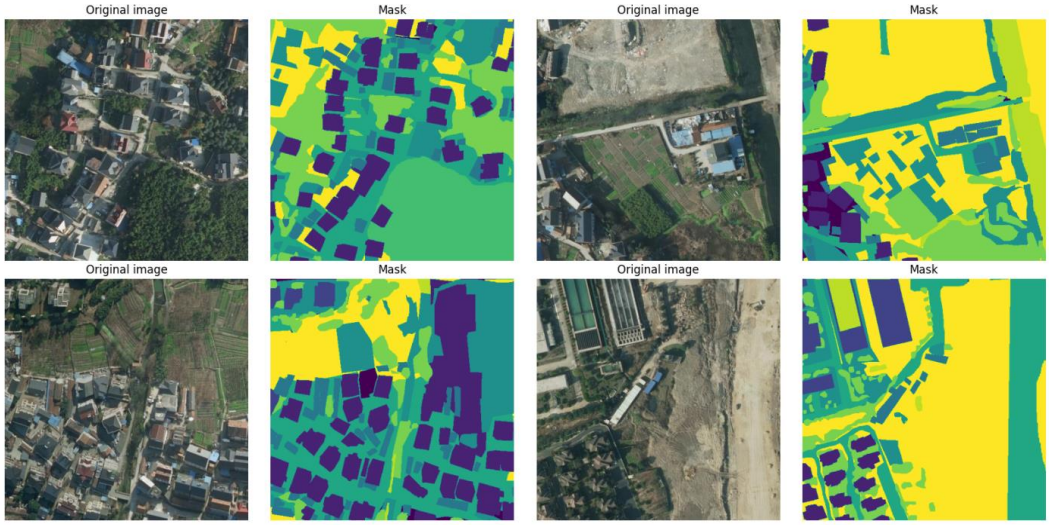

Рис. 3. Ручная аннотация с использованием Labelbox

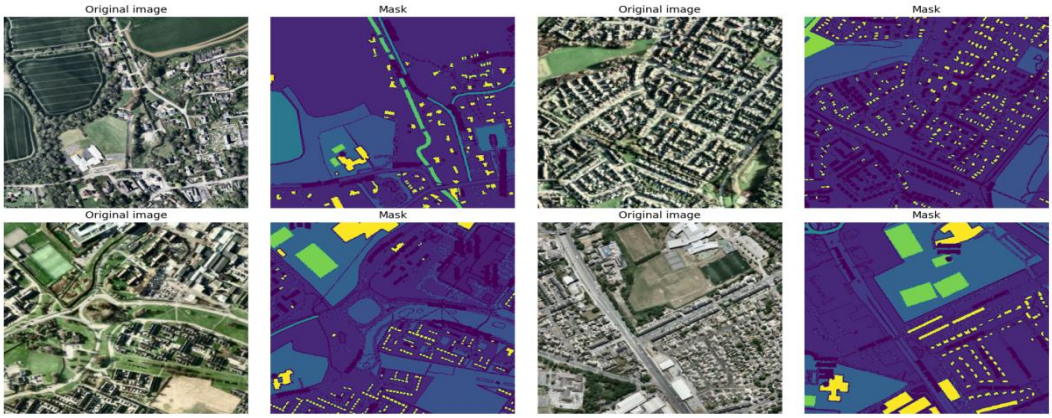

Рис. 4. Создание набора данных из общедоступной карты

В наборе данных, полученном из Google Maps, класс «растительность» был преобладающим. Чтобы устранить этот дисбаланс, сначала выбираются маски, которые содержали наибольшее количество этого класса. Для каждой выбранной маски подсчитывается количество других классов, присутствующих на том же изображении, и оценивается баланс между классами в пределах этого отдельного изображения. Затем маски были отсортированы в порядке убывания на основе количества преобладающего класса и степени дисбаланса между классами. Постепенно необходимо удалить пары изображение-маска с наихудшим балансом, постоянно отслеживая, как это повлияло на общий баланс классов в наборе данных. Это методичное сокращение помогло улучшить представление других классов без радикального уменьшения общего размера набора данных.

Гистограмма (рис. 5) показывает распределение классов в наборе данных Google Maps до и после балансировки. Значительное снижение доли

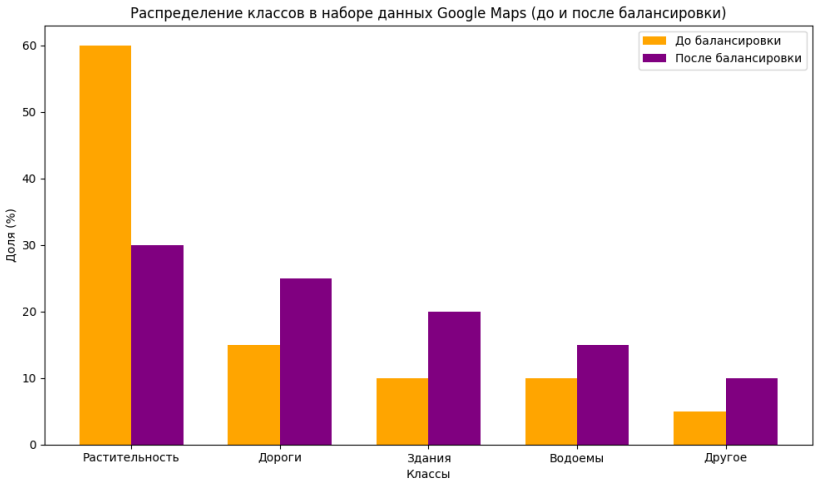

Рис. 5. Распределение классов в Google Maps

класса «растительность» (с 60% до 30%) и более равномерное распределение других классов, таких как «дороги» и «здания», подтверждают эффективность примененного подхода к устранению дисбаланса классов, что улучшает точность сегментации.

Напротив, набор данных, аннотированный в Labelbox (рис. 6), представлял более сбалансированное распределение классов. Преобладающим классом в этом наборе данных были «дороги», которые из-за своего относительного баланса остались неизменными.

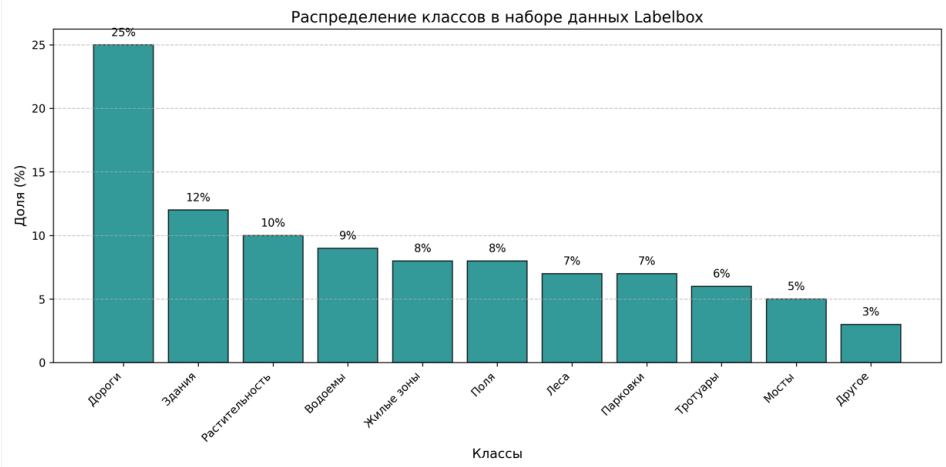

Рис. 6. Набор данных, аннотированный в Labelbox

Чтобы еще больше улучшить набор данных и увеличить изменчивость обучающих данных, были использованы методы дополнения данных, такие как повороты и отражения изображений. Этот процесс эффективно увеличил размер набора данных, предоставив моделям более разнообразные примеры и улучшив их способность обобщать невидимые дан-

ные. Необходимо соблюдать вышеописанные шаги, чтобы создать высококачественные пользовательские наборы данных, которые позволили бы выбранным моделям сегментации, в частности SegFormer, достичь высокой точности и надежности при применении к задачам сегментации спутниковых изображений.

Дисбаланс классов: наборы данных Google Maps демонстрируют преобладание растительности, что искажает обучение модели. Наборы Labelbox более сбалансированы (например, доминирующий класс — дороги). Различия в масштабах: объекты варьируются от крупных зданий до мелких деталей, что усложняет точную сегментацию на разных масштабах. Шум и вариативность: изображения из публичных карт различаются по качеству и разрешению, что вносит шум и несоответствия. Пропуски данных: не применимо, так как наборы данных тщательно отобраны, но ручные аннотации могут содержать незначительные ошибки.

4. Анализ результатов

Точная сегментация играет важную роль в ключевых областях, включая градостроительство (например, выявление границ зданий), экологический мониторинг (например, наблюдение за вырубкой лесов) и сельское хозяйство (например, анализ состояния урожая). Решение таких задач, как несбалансированность классов и масштабные различия, способствует устойчивому и надёжному функционированию систем.

График на рис. 7 демонстрирует точность модели SegFormer на различных этапах (обучение, валидация и тестирование) для шести разных наборов данных: два созданных вручную в Labelbox и четыре полученных из Google Maps. Модель показывает высокую точность на обучении практически на всех наборах, особенно на Labelbox, где достигается пик. Такое сравнение позволяет оценить влияние источника данных и качества аннотаций на способность модели обобщать и точно сегментировать новые изображения.

На графике ниже (рис. 8), представлены значения функции потерь модели SegFormer при обучении, валидации и тестировании на тех же наборах данных. Потери (loss) измеряют отклонение предсказаний модели от истинных меток и являются индикатором ошибок в сегментации.

Анализ значений потерь в дополнение к точности позволяет более полно оценить поведение модели и выявить возможные узкие места в данных.

Модель SegFormer показала высокую точность на качественно размеченных и сбалансированных наборах данных (Labelbox, Google Maps 3), где значения на обучающей, валидационной и тестовой выборках близки.

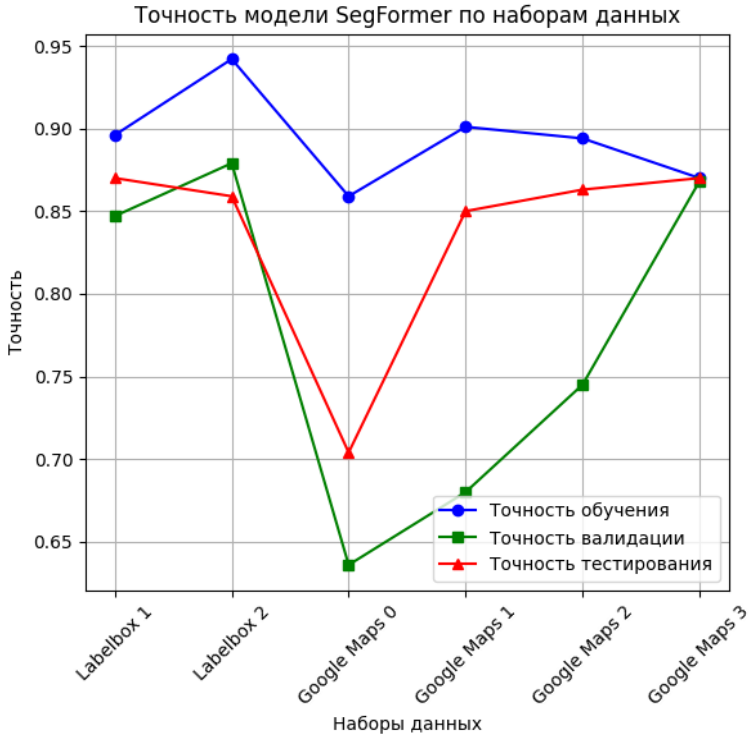

Рис. 7. Точность модели SegFormer по наборам данных

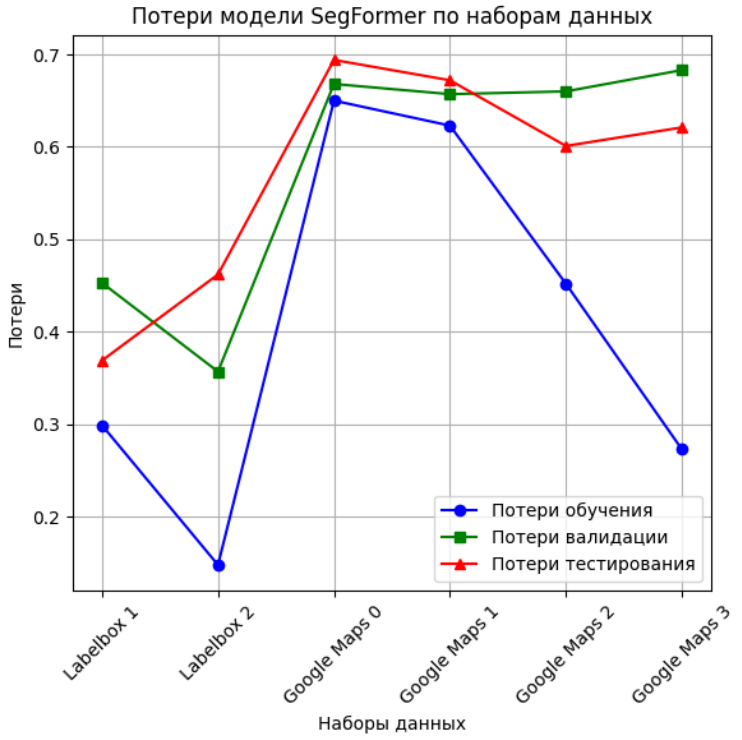

Рис. 8. Потери модели SegFormer по наборам данных

На автоматизированных наборах Google Maps 0–2 наблюдается сильное падение точности валидации и высокие потери, что связано с дисбалансом классов и шумной разметкой.

Наибольшая устойчивость результатов достигается при увеличении объёма выборки и балансировке классов, что подтверждает важность корректной подготовки данных.

Несогласованность между валидацией и тестом в отдельных случаях указывает на необходимость улучшения методики разбиения данных. Таким образом, эффективность SegFormer напрямую зависит от качества аннотации и структуры датасета, что определяет перспективность его применения при грамотной подготовке данных.

5. Заключение

В ходе исследования была успешно реализована и обучена модель семантической сегментации SegFormer для спутниковых изображений с 11 классами объектов. Тщательно подготовленные датасеты (650 пар изображений–масок для обучения, 150 – для валидации и 56 – для тестирования), сбалансированные по классам и усиленные аугментациями, позволили модели достичь на тесте точности ~0.88 и высокой mIoU, существенно превосходя классические архитектуры U-Net и DeepLabV3+. Кривые обучения и валидации продемонстрировали устойчивый рост качества и умеренное расхождение, свидетельствующее о хорошей обобщающей способности модели.

Все запланированные задачи выполнены: организована конвейерная предобработка данных, реализован собственный класс Dataset и DataLoader и механизмы борьбы с переобучением (L2-регуляризация, early stopping, чекпоинты по mIoU), а также получены количественные результаты, подтверждающие эффективность выбранного метода.

Перспективы дальнейших исследований:

- расширение наборов классов и данных;
- исследование более крупных и точных версий модели SegFormer;
- улучшение функции потерь.

Библиографические ссылки

1. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers / E. Xie [et al.] // arXiv. 2021. URL: https://arxiv.org/abs/2105.15203 (date of access: 08.05.2025).

2. *Ronneberger O., Fischer P., Brox T.* U-Net: Convolutional Networks for Biomedical Image Segmentation. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention // arXiv. 2015. URL: https://arxiv.org/abs/1505.04597 (date of access: 13.05.2025).

3. *Chen L. C., Papandreou G., Schroff F.* Rethinking Atrous Convolution for Semantic Image Segmentation // arXiv. 2017. URL: https://arxiv.org/abs/1706.05587 (date of access: 14.05.2025).

4. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale / A. Dosovitskiy [et al.] // arXiv. 2020. URL: https://arxiv.org/abs/2010.11929 (date of access: 16.05.2025).

5. *Liu Z., Lin Y., Cao Y.* Swin Transformer: Hierarchical Vision Transformer using Shifted Windows // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. P. 10012–10022. URL: https://ieeexplore.ieee.org/document/9710580 (date of access: 23.05.2025).

6. *Long J., Shelhamer E., Darrell T.* Fully Convolutional Networks for Semantic Segmentation // arXiv. 2015. URL: https://arxiv.org/pdf/1505.04597 (date of access: 19.06.2025).