

## **ИССЛЕДОВАНИЕ И АНАЛИЗ СОВРЕМЕННЫХ МЕТОДОВ ОЦИФРОВКИ АРХИВНЫХ ГЕОЛОГИЧЕСКИХ КАРТ**

Б.И. Григорьев<sup>1</sup>, С.А. Сазонова<sup>1</sup>

<sup>1</sup>ФГБОУ ВО «Воронежский государственный лесотехнический университет  
имени Г.Ф. Морозова»

Аннотация. В работе рассматриваются основные методы оцифровки архивных геологических карт, их основные преимущества и недостатки. Описывается процесс обработки изображений геологической карты каждого метода оцифровки.

Ключевые слова: геологическая карта, ручной метод оцифровки, полуавтоматический метод оцифровки, автоматический метод оцифровки, векторизация, геоинформационные системы, нейронная сеть.

## **RESEARCH AND ANALYSIS OF MODERN METHODS OF DIGITIZATION OF GEOLOGICAL MAPS**

B.I. Grigorev<sup>1</sup>, S.A. Sazonova<sup>1</sup>

<sup>1</sup>Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. The paper discusses the main methods of digitizing archival geological maps, their main advantages and disadvantages. The process of processing images of the geological map of each digitization method is described.

Keywords: geological map, manual digitization method, semi-automatic digitization method, automatic digitization method, vectorization, geoinformation systems, neural network.

С каждым днем объем геологических данных стремительно растет, в том числе за счет сохраненных в архивах старых геологических карт. Эти карты содержат ценную информацию о геологической структуре и ресурсах земной коры, однако их использование и анализ требуют трудоемких процессов оцифровки и интерпретации.

Оцифровка любого картографического материала, в частности архивных геологических карт, состоит из следующих этапов:

1. Наличие исходных данных. Основой для создания цифрового картографического материала (ЦКМ) служат старые архивные материалы с изображением топографического плана. Это могут быть планшеты на жесткой основе (фанера или алюминиевые пластины), на кальке или лавсановой плёнке вычерченные тушью, а также бумажные копии выполненные в электронном виде.

2. Сканирование. Для того чтобы перевести в электронный вид (оцифровать) картографический материал, используется сканер или цифровая фотокамера с высоким разрешением. В процессе сканирования (фотографирования) получается массив точек в виде растрового изображения, который будет служить основой (подложкой) для создания ЦКМ.

3. Привязка растрового изображения. Процесс, при котором растровое изображение соотносится с географическими координатами.

4. Обработка растрового изображения. Все действия, производимые на данном этапе, направлены на улучшение изображения, сохранение детализации и подготовку масок (слоев с однотипной информацией) для распознавания на них объектов

5. Векторизация. Это процесс преобразования растровых изображений в векторное представление, использующее набор кривых и точек для более точного описания изображения. В процессе векторизации к каждому объекту, оцифрованному на карте, добавляются соответствующие атрибуты (название, классификация, значения и т. д.); присваиваются географические координаты объектам на карте для определения их местоположения на земной поверхности.

6. Корректировка. В завершающем этапе производится анализ на соответствие реальным географическим данным и коррекция возможных ошибок.

Основные методы векторизации картографического материала в геоинформационных системах (ГИС)

Ручной метод оцифровки на сегодняшний день можно отнести к традиционным методам (наряду со сканированием), но несмотря на появление новых методик, данный метод не потерял своей актуальности. Суть данного метода заключается в том, что каждый этап оцифровки от сканирования до векторизации и экспорта в нужный формат осуществляется с участием человека (оператора). Стоит отдельно пояснить некоторые этапы оцифровки в ручном режиме.

Импорт растрового изображения после его сканирования осуществляется вручную в специальную программу для векторизации.

Привязка растрового изображения геологической карты осуществляется путем выбора нужной системы координат и привязкой углов карты и её центральной точки. Геологическая карта чаще всего уже разбита сеткой координат с указанием долготы и широты, а также специальной номенклатурой, кодировка которой указывает её территориальное расположение.

На бумажных носителях часто встречаются дефекты в виде потертостей, пятен и других дефектов, затрудняющих автоматическое распознавание. Для улучшения изображений применяются различного рода встроенные в ПО фильтры (фильтр увеличения резкости, бикубическая интерполяция, гистограммное выравнивание и др.), которые оператор применяет по своему усмотрению.

В общем случае векторизация растрового изображения в ручном режиме осуществляется посредством расстановки точек или отрезков на карте, совмещая их с видимыми объектами.

Достоинства ручного метода оцифровки:

- высокое качество: возможность более тщательной и точной обработки изображений;
- гибкость и креативность: возможность внесения различных коррекций и улучшений в процессе обработки;
- сохранение деталей: способность сохранять детали и особенности изображения при оцифровке;
- контроль над процессом: возможность контроля каждого этапа оцифровки.

Недостатки ручного метода оцифровки:

- времязатратность: требуется больше времени на обработку каждого изображения;
- высокая стоимость: затраты на оплату труда специалистов, занимающихся ручной оцифровкой;
- неэффективность для больших объемов: при большом количестве материалов ручной метод может быть неэффективен и неудобен.

Полуавтоматический метод оцифровки – процесс, который включает человеческое вмешательство для помощи системе в распознавании и преобразовании данных. Вот общий процесс полуавтоматической оцифровки:

- предобработка данных: подготовленные изображения или данные проходят предварительную обработку, включающую коррекцию искажений, устранение шума и другие техники для улучшения качества.

– Интерактивная сегментация: пользователь или оператор взаимодействует с системой для пометки ключевых областей, объектов или контуров на изображении, помогая системе определить структуру данных.

– Частичное распознавание: система использует комбинацию алгоритмов автоматического распознавания и пользовательских меток для выделения и идентификации объектов на изображении.

– Ручная коррекция и добавление информации: пользователь вручную корректирует и дополняет обработанные данные (атрибуты) там, где системе сложно правильно распознать или векторизовать объекты.

– При полуавтоматической векторизации или интерактивной – часть операций производится автоматически. Так, например, при векторизации горизонталей достаточно задать начальную точку и направление отслеживания линий, далее векторизатор сам отследит эту линию до тех пор, пока на его пути не встретятся неопределенные ситуации, типа разрыва линии. Возможности интерактивной векторизации прямо связаны с качеством исходного материала и сложностью карты

Достоинства полуавтоматического метода оцифровки:

– высокая точность: участие человека позволяет исправить ошибки, которые могут возникнуть в процессе автоматической обработки, повышая общую точность оцифровки;

– гибкость и адаптивность: человеческое вмешательство позволяет легко адаптировать процесс оцифровки под конкретные требования и особенности данных;

– экономия времени: автоматические методы могут значительно сократить время обработки, и человеческое вмешательство используется там, где это действительно необходимо;

– улучшение качества данных: при полуавтоматическом методе возможно проведение дополнительных проверок и исправлений, что способствует повышению качества оцифрованных данных.

Недостатки полуавтоматического метода оцифровки:

– затраты на человеческий ресурс: участие человека в процессе оцифровки может потребовать значительных затрат времени и ресурсов;

– субъективность и ошибки человеческого фактора: вмешательство человека может привести к человеческим ошибкам или субъективным оценкам, что может повлиять на качество и объективность оцифрованных данных;

– сложность масштабирования: полуавтоматический метод может оказаться сложным для масштабирования, особенно если требуется обработка большого объема данных;

– высокая стоимость: использование человеческого ресурса для дополнительной проверки и исправления данных может сделать процесс оцифровки более затратным.

Автоматический метод оцифровки в настоящий момент включает в себя использование специализированных геоинформационных систем (ГИС) и программного обеспечения для преобразования данных с геологических карт в цифровой формат.

Автоматическая векторизация чаще всего применяется на черно-белых геологических картах (бинарные изображения). В данном случае программа-векторизатор распознает черный цвет и переводит его в векторный формат. Белый цвет распознается, как фон. Для этого может быть использован алгоритм Стэджера [1] для уменьшения времени временных затрат. Для восстановления возможных разрывов при распознавании изолиний используется интерполяция кривой Безье по трем точкам [1].

Обработку растрового изображения в автоматическом режиме можно разделить на два этапа: предобработка изображения и распознавание интересующих объектов на изображении (в нашем случае векторизация изолиний). Каждый этап, в свою очередь, включает в себя ряд подэтапов. Так, для предобработки изображения сначала применяется кластеризация по признаку цветности, затем скелетизация и морфологические операции для восстановления разрывов у изолиний, образовавшихся на предыдущих этапах предобработки или из-за плохого качества изображения.

Привязка раstra в ГИС-программе означает соотнесение пиксельных координат на изображении с реальными географическими координатами на карте. Этот процесс основан на поиске соответствия между пикселями на изображении и их географическими координатами в градусах или метрах на карте. Таким образом, программа понимает, как точки на растровом изображении соотносятся с конкретными местами на карте. В основе привязки растрового изображения лежат два алгоритма: алгоритм трансформации раstra, изменяющий изображение таким образом, чтобы оно совпадало с имеющимися географическими координатами, и алгоритм «ресемплинг» или передискретизация, в основе которого лежит изменение разрешения изображения для корректного сопоставления с местностью.

Достоинства автоматического метода оцифровки:

- Скорость: автоматический метод оцифровки данных позволяет обрабатывать большие объемы информации значительно быстрее, чем это делалось бы вручную.
- Эффективность: автоматизированные процессы обработки данных могут повысить эффективность работы за счет минимизации человеческого вмешательства.
- Масштабируемость: автоматические методы оцифровки легче масштабируются для работы с большими объемами данных и их быстрой обработки.
- снижение ошибок: при правильной настройке и калибровке системы автоматической обработки, можно уменьшить вероятность ошибок по сравнению с человеческим вмешательством.

Недостатки автоматического метода оцифровки:

- недостаточная точность: автоматические методы могут быть менее точными, особенно при работе с нестандартными или плохо структурированными данными;
- ограниченная гибкость: в случае изменения требований или особенностей данных, автоматизированные системы могут показать себя менее гибкими по сравнению с человеческим вмешательством
- необходимость настройки: для достижения оптимальной производительности и точности автоматизированной системы оцифровки требуется тщательная настройка и обучение;
- трудность обработки специфических случаев: в случае нестандартных данных или особых требований автоматический метод может столкнуться с трудностями, которые человеческий фактор мог бы легче преодолеть.

Существует еще один метод оцифровки архивных геологических карт – с применением машинного обучения. В основе данного метода лежит создание и обучение нейронной сети. Нейронная сеть — это математическая модель и ее реализация в виде программной или программно-аппаратной реализации, которая основана на моделировании активности биологических нейронных сетей, которые представляют собой сети нейронов в биологическом организме. [2] Каждый нейрон принимает входные данные, обрабатывает их и передает результат следующему нейрону. Нейроны организованы в слои, и информация передается от входного слоя к выходному слою через промежуточные слои, называемые скрытыми слоями.

Обучение нейронной сети - процесс настройки параметров нейронной сети на основе обучающих данных с целью минимизации ошибки или потерь во время выполнения конкретной задачи. Обучение нейросетей происходит в два этапа:

1. прямое распространение, при котором нейросеть в тестовом режиме «прогоняет» через себя данные и прогнозирует результат;
2. обратное распространение ошибки, при котором погрешности каждого звена отправляются обратно в виде градиента, на основании чего изменяются веса. [3]

Существуют три основных способа обучения нейросетей:

1. обучение с учителем – данный способ обучения предполагает наличие обучающих данных, где для каждого входа имеется соответствующий правильный выход.
2. Обучение без учителя - способ предполагает спонтанный вид самообучения, в котором нет размеченных данных. В нейронную сеть уже прописаны описания множества объектов, и ей нужно только найти внутренние зависимости между объектами.
3. Обучение с подкреплением. В обучении с подкреплением нейронная сеть учится на основе взаимодействия с окружающей средой. Сеть принимает действия в данной среде и получает обратную связь в виде награды или штрафа, что позволяет ей корректировать свое поведение для достижения целей.

Рассмотрим стандартные задачи, решаемые нейронными сетями в приложении к изображениям [4]:

- идентификация объектов;
- распознавание частей объектов (например, лиц, рук, ног и т.д.);
- семантическое определение границ объектов (позволяет оставлять только границы объектов на картинке);
- семантическая сегментация (позволяет разделять изображение на различные отдельные объекты);
- выделение нормалей к поверхности (позволяет преобразовывать двумерные картинки в трехмерные изображения);
- выделение объектов внимания (позволяет определять то, на что обратил бы внимание человек на данном изображении).

С учетом решения успешных кейсов в ГИС с использованием машинного обучения [5], а также в других областях [6], выделим основные достоинства и недостатки данного метода:

- автоматизация: использование нейросетевых алгоритмов и технологии компьютерного зрения позволяет автоматизировать процесс оцифровки геологических карт, что увеличивает производительность и снижает трудозатраты.
- скорость: нейросетевые алгоритмы могут обрабатывать большие объемы данных быстрее, чем традиционные методы, что способствует более быстрой оцифровке геологических карт;
- высокая точность: при правильной настройке и обучении нейронных сетей, можно добиться высокой точности при распознавании и оцифровке геологических объектов на картах;
- способность обрабатывать сложные структуры: нейросетевые алгоритмы и технология компьютерного зрения способны обрабатывать сложные геологические структуры и образования, что может быть сложно для человека.

Недостатки применения нейросетевых алгоритмов и технологии компьютерного зрения для оцифровки геологических карт:

- требование большого объема обучающих данных: для достижения высокой точности нейросетевые алгоритмы требуют обширные обучающие наборы данных, что может потребовать значительных усилий и ресурсов.
- сложность в интерпретации результатов: иногда сложно понять, как нейросеть приняла свое решение, что может затруднить валидацию результатов оцифровки;
- необходимость корректировки и улучшения моделей: нейросетевые модели требуют постоянного обновления, корректировки и улучшения для обеспечения высокой точности и актуальности результатов оцифровки;
- ограничения в обработке специфических случаев: нейросетевые алгоритмы могут столкнуться с трудностями при обработке нестандартных геологических карт или объектов, которые не входили в обучающий набор данных.

### Заключение

Использование нейросетевых алгоритмов и технологии компьютерного зрения для оцифровки геологических карт предоставляет значительные преимущества в виде автоматизации, скорости, точности и способности обработки сложных структур. Однако данные методы требуют большого объема обучающих данных и постоянной настройки моделей для достижения высокой эффективности. Поэтому выбор между традиционными методами и нейросетевыми алгоритмами зависит от конкретных потребностей проекта и доступных ресурсов.



## Список литературы

1. Курочкин, В. Ю. Разработка сервиса для автоматизированного поиска и распознавания изолиний на изображениях геологических карт / В. Ю. Курочкина, А. А. Ступников. — Текст : электронный // Математическое и информационное моделирование : материалы Всероссийской конференции молодых ученых (Тюмень, 18–20 мая 2023 г.) / Министерство науки и высшего образования РФ, Тюменский государственный университет, Институт математики и компьютерных наук ; редакционная коллегия : Е. П. Вдовин [и др.]. — Тюмень : ТюмГУ-Press, 2023. — Вып. 21. — С. 144–152.
2. Никитин, А. А. Процесс распознавания изображения нейронной сетью / А. А. Никитин, Н. И. Лиманова. — Текст : непосредственный // Молодой ученый. — 2020. — № 47 (337). — С. 23-25. — URL: <https://moluch.ru/archive/337/75420/> (дата обращения: 27.03.2024).
3. Левченко, К. М. Нейронные сети = Neural networks / Левченко К. М., Сыч А. А. // Научная конференция учащихся колледжа : материалы 58-й научной конференции аспирантов, магистрантов и студентов БГУИР, Минск, 18–22 апреля 2022 / Белорусский государственный университет информатики и радиоэлектроники, Минский радиотехнический колледж ; редкол.: В. В. Шаталова [и др.]. — Минск : БГУИР, 2022. — С. 89–93.
4. Маркова, С. В. Применение нейронной сети для создания системы распознавания изображений / С. В. Маркова, К. Ю. Жигалов // Фундаментальные исследования. — 2017. — № 8-1. — С. 60-64. — EDN ZELKMR.
5. Селлин, А. Ю. Актуализация геоинформационных систем посредством интеграции нейронных сетей / А. Ю. Селлин, О. Н. Понамарева // Состояние и перспективы развития современной науки по направлению «Геоинформационные платформы военного назначения»: Сборник статей I Всероссийской научно-технической конференции, Анапа, 17 марта 2021 года. — Анапа: Федеральное государственное автономное учреждение "Военный инновационный технополис "ЭРА", 2021. — С. 195-205. — EDN GHCCMY.
6. Создание поведенческой модели LDMOS транзистора на основе искусственной MLP нейросети и ее описание на языке Verilog-A / С.А. Победа, М.И. Черных, Ф.В. Макаренко, К.В. Зольников // Моделирование систем и процессов. — 2021. — Т. 14, № 2. — С. 28-34. — DOI: 10.12737/2219-0767-2021-14-2-28-34.

7. Полуэктов А.В., Макаренко Ф.В., Ягодкин А.С. Использование сторонних библиотек при написании программ для обработки статистических данных // Моделирование систем и процессов. – 2022. – Т. 15, № 2. – С. 33-41.

### References

1. Kurochkin, V. Yu. Development of a service for automated search and recognition of isolines on images of geological maps / V. Yu. Kurochkina, A. A. Stupnikov. — Text : electronic // Mathematical and information modeling : materials of the All-Russian Conference of Young Scientists (Tyumen, May 18-20, 2023)

2. Nikitin, A. A. The process of image recognition by a neural network / A. A. Nikitin, N. I. Limanova. — Text : direct // Young scientist. — 2020. — № 47 (337). — Pp. 23-25. — URL: <https://moluch.ru/archive/337/75420/> / (date of access: 03/27/2024).

3. Levchenko, K. M. Neural networks = Neural networks / Levchenko K. M., Sych A. A. // Scientific conference of college students : materials of the 58th scientific conference of graduate students, undergraduates and students of BSUIR, Minsk, April 18-22, 2022 / Belarusian State University of Informatics and Radioelectronics, Minsk Radio Engineering College ; editorial board: V. V. Shatalova [et al.]. – Minsk : BGUIR, 2022. – pp. 89-93.

4. Markova, S. V. Application of a neural network to create an image recognition system / S. V. Markova, K. Yu. Zhigalov // Fundamental research. - 2017. – No. 8-1. – pp. 60-64. – EDN ZELKMR.

5. Sellin, A. Y. Actualization of geoinformation systems through the integration of neural networks / A. Y. Sellin, O. N. Ponamareva // The state and prospects of development of modern science in the field of "Military geoinformation platforms": Collection of articles of the I All-Russian Scientific and Technical Conference, Anapa, March 17, 2021. – Anapa: Federal State Autonomous Institution "Military Innovative Technopolis "ERA", 2021. – pp. 195-205. – EDN GHCCMY.

6. Creation of a behavioral model of an LDMOS transistor based on an artificial MLP neural network and its description in the Verilog-A language / S.A. Pobeda, M.I. Chernykh, F.V. Makarenko, K.V. Zolnikov // Modeling of systems and processes. - 2021. – Vol. 14, No. 2. – pp. 28-34. – DOI: 10.12737/2219-0767-2021-14-2-28-34.

7. Poluektov A.V., Makarenko F.V., Yagodkin A.S. The use of third-party libraries when writing programs for processing statistical data // Modeling of systems and processes. - 2022. – Vol. 15, No. 2. – pp. 33-41.