

В. А. Зуев^{1✉}, П. Ю. Бугаков¹

Применение мультимодальных нейронных сетей для автоматизации оценки визуальной информативности трехмерных карт городских территорий

¹Сибирский государственный университет геосистем и технологий,
г. Новосибирск, Российская Федерация
e-mail: veuzdalv638@gmail.com

Аннотация. В статье исследуется возможность автоматизации оценки визуальной информативности трехмерных карт городских территорий с использованием мультимодальных нейронных сетей. Для проведения эксперимента были использованы три трехмерные карты городских территории с разной плотностью застройки и рельефом. Для оценки визуальной информативности трехмерных карт был предложен набор из 14 параметров, включающий в себя количество видимых объектов, цветовой контраст, детализация объектов, размер условных знаков и др. Определение значений каждого из параметров выполнялось с использованием мультимодальных нейронных сетей Eagle 2, Qwen 2.5, Llama 4, ChatGPT 4o, Gemma 3. Полученные значения были использованы для расчета комплексных показателей визуальной информативности анализируемых трехмерных карт. Точность оценки визуальной информативности трехмерных карт мультимодальными нейронными сетями определялась путем вычисления отклонения расчетных показателей от контрольных значений, полученных путем экспертной оценки. Результаты исследования демонстрируют потенциал мультимодальных нейронных сетей для автоматизации оценки визуальной информативности трехмерных карт, а также выявляют особенности применения различных моделей для решения данной задачи.

Ключевые слова: мультимодальные нейронные сети, трехмерные карты, визуальная информативность, автоматизированная оценка, городская среда

V. A. Zuev^{1✉}, P. Y. Bugakov¹

Application of multimodal neural networks for automation of visual informativeness assessment of three-dimensional maps of urban areas

¹Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation
e-mail: veuzdalv638@gmail.com

Abstract. The article explores the possibility of automating the assessment of the visual informativeness of three-dimensional maps of urban areas using multimodal neural networks. To conduct the experiment, three three-dimensional maps of urban areas with different building densities and terrain were used. To assess the visual informativeness of three-dimensional maps, a set of 14 parameters was proposed, including the number of visible objects, color contrast, object detail, the size of conventional signs, etc. The values of each parameter were determined using multimodal neural networks Eagle 2, Qwen 2.5, Llama 4, ChatGPT 4o, Gemma 3. The obtained values were used to calculate complex indicators of the visual informativeness of the analyzed three-dimensional maps. The accuracy of the assessment of the visual information content of three-dimensional maps by multimodal neural networks was determined by calculating the deviation of the calculated indicators from the control values obtained by expert evaluation. The results of the study demonstrate the potential of

multimodal neural networks for automating the assessment of the visual information content of three-dimensional maps, as well as identify the features of using various models to solve this problem.

Keywords: multimodal neural networks, three-dimensional maps, visual informativeness, automated assessment, urban environment

Введение

Современные технологии машинного зрения и искусственного интеллекта открывают новые возможности для автоматизации анализа сложных визуальных объектов, к которым можно отнести трехмерные карты городских территорий. На сегодняшний день трехмерные карты нашли широкое применение в различных областях деятельности человека, таких как: городское планирование, навигация, туризм, управление инфраструктурой [1]. Эффективность использования трехмерных карт напрямую зависит от их визуальной информативности, которая определяется как отношение количества воспринимаемой человеком визуальной информации о геопространстве к общему количеству информации, заложенной в карте ее составителем [4].

Традиционные методы оценки визуальной информативности трехмерных карт являются трудоемкими, субъективными и требуют экспертных знаний. Автоматизация этого процесса с использованием мультимодальных нейронных сетей (ММНС), может значительно повысить эффективность и объективность оценки.

Целью данного исследования является оценка возможности применения ММНС для автоматизированной оценки визуальной информативности трехмерных карт городских территорий. Для достижения этой цели были поставлены следующие задачи:

- определить ключевые параметры, характеризующие визуальную информативность трехмерных карт;
- разработать план эксперимента по оценке визуальной информативности трехмерных карт с использованием ММНС;
- выполнить сравнительный анализ результатов, полученных с использованием различных ММНС.

Мультимодальные нейронные сети

Мультимодальные нейронные сети (ММНС) – это класс нейронных сетей, способных одновременно обрабатывать и интегрировать данные различных типов (модальностей), таких как текст, изображения, звук и числовые параметры [2]. В отличие от традиционных нейронных сетей, которые работают с данными одного типа, ММНС позволяют учитывать взаимосвязи между различными модальностями. Мультимодальные нейронные сети делятся по типу слияния модальностей, существуют модели с ранним слиянием (early fusion), поздним слиянием (late fusion), а также с гибридным слиянием (hibrid fusion) [3]. Модели с ранним слиянием показывают высокую эффективность в малых нейронных сетях, с поздним слиянием показывают наилучший результат в больших.

В последнее время наблюдается значительный прогресс в области ММНС, что привело к появлению новых моделей с открытым и закрытым исходным кодом. Примерами таких моделей являются Gemma 3, ChatGPT-4o, Llama 4 и др. Однако применение этих моделей для оценки визуальной информативности трехмерных карт требует адаптации и оценки их эффективности в контексте данной задачи.

Методы и материалы

В качестве данных для исследования были использованы три трехмерные карты городских территорий. В основе карт лежат модели различных городских районов с разной плотностью застройки и рельефом. Характеристики карт приведены в табл. 1, трехмерные карты представлены на рис.1–3.

Таблица 1

Характеристики карт

Карта №	Детализация (LOD)	Тематика	Площадь (км ²)
1	LOD1	Жилой район	1
2	LOD2	Набережная	7
3	LOD3	Центр города	2

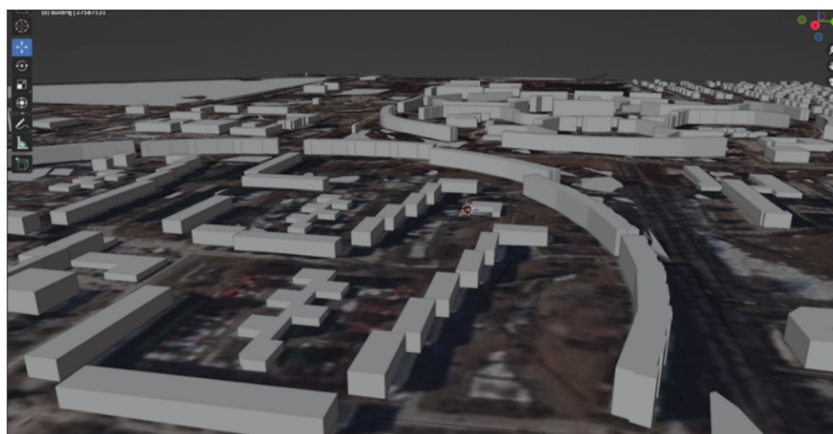


Рис. 1. Карта №1 – рабочий поселок Краснообск



Рис. 2. Карта №2 – город Хабаровск



Рис. 3. Карта №3 – город Манчестер

Для оценки визуальной информативности трехмерных карт были выбраны следующие параметры [4]:

- количество видимых объектов – целочисленный параметр;
- цветовой контраст (коэффициент 1–21) – соответствие стандартам WCAG (AA/AAA) [5];
- детализация объектов – уровни LOD0–LOD4 (1–5 баллов);
- размер условных знаков – отношение площади знаков к площади карты (%);
- количество условных знаков – целочисленный параметр;
- качество текстур – плотность текселей (0.1–10.24 px/cm);
- читаемость – шкала 1–5 (от "очень низкой" до "отличной");
- угол наклона – 0°–90°;
- поле зрения – 30°–150°;
- изрезанность рельефа – 1 (резкопересеченный) – 3 (равнинный);
- плотность застройки – 1 (редкая) – 3 (густая);
- этажность застройки – 1 (малоэтажная) – 4 (многоэтажная);
- распределение объектов – 1 (рассредоточенное) – 3 (равномерное);
- средняя площадь элемента – в м².

План эксперимента

Для оценки визуальной информативности трехмерных карт с использованием ММНС был составлен следующий порядок действий:

1. Для каждой ММНС составляется промт следующего содержания: «Оцените визуальную информативность представленной трехмерной карты городских территорий по следующим параметрам: [перечисление параметров, представленных выше]. Предоставьте значения для каждого параметра в соответствующих единицах измерения».
2. Каждой ММНС предоставляются анализируемые трехмерные карты.
3. Значения параметров, полученные от каждой ММНС, записываются и используются для расчета комплексного показателя визуальной информативности анализируемой карты.

4. Вычисляются отклонения рассчитанных показателей от контрольных значений, полученных путем экспертной оценки.

Расчет комплексного показателя визуальной информативности

Комплексный показатель визуальной информативности трехмерной карты рассчитывался как взвешенная сумма оценок качества отдельных параметров по формуле (1).

$$K = \sum_{i=1}^N g_i \cdot k_i, \quad (1)$$

где K – комплексный показатель визуальной информативности трехмерной карты; N – количество учитываемых свойств; g_i – коэффициент весомости i -го свойства; k_i – показатель качества, соотнесенный с нормативами.

Веса параметров были определены методом рангов на основе экспертных оценок.

Для расчета оценки качества каждой карты использовалась формула (2).

$$k_i = \frac{p_i - p_i^{\min}}{p_i^{\max} - p_i^{\min}}, \quad (2)$$

где k_i – показатель качества, соотнесенный с нормативами; p_i – измеренная величина показателя качества на исследуемой карте; $p_i^{\max}$, $p_i^{\min}$ – нормативные значения, за пределами которых карту невозможно или нецелесообразно использовать. Для показателей, которых нет в нормативах, данные показатели вычисляются путём просмотра и оценки множества трехмерных картографических изображений.

Результаты

В ходе эксперимента были использованы следующие мультимодальные нейронные сети: Eagle 2, Qwen 2.5, Llama 4, ChatGPT 4o и Gemma 3. Janus Pro и GigaChat 2 Max были исключены из дальнейшего анализа из-за неудовлетворительных результатов, так как Janus Pro не предоставил никаких оценок, а GigaChat 2 Max предоставил одинаковые оценки для всех карт. Результаты оценки визуальной информативности трехмерных карт с использованием различных ММНС представлены в табл. 2.

Таблица 2

Результаты оценки визуальной информативности трехмерных карт
с использованием ММНС

Модель	Карта №1	Карта №2	Карта №3	Среднее время оценки (сек)
Eagle 2	276.5	444.82	576.6	81
Qwen 2.5	364.34	587.35	698.57	153
Llama 4	368.26	540.36	639.86	114
ChatGPT 4o	293.14	483.57	605.75	75
Gemma 3	276.23	453.26	624.7	97

Значения в табл. 2 представляют собой общую оценку визуальной информативности карты, рассчитанную по формуле (1) с учетом весов параметров, умноженную на 1000.

В табл. 3 представлены отклонения от контрольных значений, полученных путем экспертной оценки.

Таблица 3

Отклонение от контрольных значений

Модель	Карта №1	Карта №2	Карта №3	Общая Дельта
Eagle 2	19.16	17.24	72.87	109.27
Qwen 2.5	106.99	125.29	49.1	281.38
Llama 4	110.91	78.3	9.61	198.82
ChatGPT 4o	35.8	21.52	43.73	101.05
Gemma 3	18.89	8.79	24.78	52.46

Обсуждение

Результаты исследования показали, что мультимодальные нейронные сети могут быть успешно использованы для автоматизированной оценки визуальной информативности трехмерных карт городских территорий. Модель Gemma 3 продемонстрировала наименьшее отклонение от экспертной оценки, что свидетельствует о ее высокой точности и надежности. Модели Eagle 2, Qwen 2.5 и Llama 4 показали более значительные отклонения, что может быть связано с особенностями их архитектуры и обучения.

Необходимо отметить, что точность оценки визуальной информативности трехмерных карт зависит от множества факторов, включая качество данных, выбор параметров оценки и весов параметров. Дальнейшие исследования должны быть направлены на оптимизацию этих факторов, а также проведение дальнейшего обучения предварительно обученной модели на более мелких специализированных датасетах с целью совершенствования их возможностей и повышения точности в конкретной задаче.

Заключение

В данной статье была исследована возможность применения мультимодальных нейронных сетей для автоматизированной оценки визуальной информативности трехмерных карт городских территорий. Результаты исследования показали, что ММНС могут быть успешно использованы для решения данной задачи, а модель Gemma 3 продемонстрировала наилучшие результаты. Полученные результаты могут стать основой для дальнейшей разработки автоматизированной системы оценки визуальной информативности трехмерных карт, что позволит повысить скорость и объективность процесса оценки.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Трехмерная компьютерная картография [Текст] : монография / Д. В. Лисицкий, П. Ю. Бугаков, Нгуен Ань Тай ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего образования "Сибирский государственный университет геосистем и технологий" (СГУГиТ). – Новосибирск : СГУГиТ, 2016. — 178 с.
2. Ghosh and K. Deepa, "QueryMintAI: Multipurpose Multimodal Large Language Models for Personal Data," in IEEE Access, vol. 12, pp. 144631-144651, 2024, doi: 10.1109/ACCESS.2024.3468996.
3. Морошкин, Н. А. Обзор современных мультимодальных архитектур нейросетей / Н. А. Морошкин // Новые информационные технологии в научных исследованиях : Материалы XXVIII Всероссийской научно-технической конференции студентов, молодых ученых и специалистов, Рязанский государственный радиотехнический университет имени В.Ф. Уткина, 22–24 ноября 2023 года. – Рязань: Рязанский государственный радиотехнический университет имени В.Ф.Уткина, 2023. – С. 108–111.
4. Зуев, В. А. Факторы, влияющие на визуальную информативность трехмерных карт / В. А. Зуев, П. Ю. Бугаков // Аллея науки. – 2023. – Т. 1, № 6(81). – С. 1557–1560.
5. Understanding WCAG 2.0: Visual and Audio Contrast [Электронный ресурс] // W3C Web Accessibility Initiative. – URL: <https://www.w3.org/TR/UNDERSTANDING-WCAG20/visual-audio-contrast-contrast.html> (дата обращения: 21.04.2025).

© В. А. Зуев, П. Ю. Бугаков, 2025