

*Касим Роман Хайдарович, бакалавр,
МГТУ им. Н.Э. Баумана,
г. Москва,*

АППАРАТНЫЕ И ПРОГРАММНЫЕ МЕТОДЫ ОТСЛЕЖИВАНИЯ ВЗГЛЯДА

***Аннотация:** В статье рассматриваются основные методы отслеживания взгляда человека. Проведено сравнение аппаратного подхода, основанного на применении специализированных инфракрасных айтрекеров, и программного подхода, использующего стандартные камеры и алгоритмы компьютерного зрения. Рассмотрены принципы работы, преимущества, ограничения и области применения каждого подхода. Отдельное внимание уделено нейросетевым методам оценки направления взгляда. Представлен обзор архитектур iTracker, Special Weights CNN и L2CS-Net, а также выполнено их сравнение по средней угловой и средней позиционной ошибкам.*

***Ключевые слова:** отслеживание взгляда, айтрекинг, компьютерное зрение, глубокое обучение, свёрточные нейронные сети.*

HARDWARE- AND SOFTWARE-BASED METHODS FOR EYE TRACKING

***Abstract:** The article is devoted to the consideration of the main methods used for human eye tracking. A comparison is made between the hardware-based approach, which relies on specialized infrared eye trackers, and the software-based approach, which uses conventional cameras and computer vision algorithms. The operating principles, advantages, limitations, and application*

areas of both approaches are discussed. Particular attention is paid to neural network methods for gaze estimation. The iTracker, Special Weights CNN, and L2CS-Net architectures are reviewed and compared in terms of mean angular error and mean positional error.

Key Words: *eye tracking, gaze estimation, computer vision, deep learning, convolutional neural networks.*

Введение

Отслеживание взгляда позволяет получать данные об особенностях зрительного внимания и поведения человека, включая процессы принятия решений, распределение внимания, уровень усталости и когнитивную нагрузку. Технологии отслеживания взгляда применяются в научных и медицинских исследованиях, маркетинге, образовании, UX-аналитике, а также в системах управления устройствами.

Для определения направления взгляда используются аппаратные и программные методы, различающиеся принципом работы, точностью, стоимостью и условиями применения. Отдельное направление развития технологии связано с применением нейронных сетей, позволяющих оценивать направление взгляда по изображениям лица и глаз, полученным с помощью обычных камер.

Методы отслеживания взгляда

В настоящее время существует два основных подхода к решению задачи отслеживания взгляда: аппаратный, основанный на использовании инфракрасной оптики, и программный, где применяются алгоритмы компьютерного зрения для анализа изображений глаз [1].

Аппаратный подход предполагает применение специализированных устройств, называемых айтрекерами, принцип работы которых основан на подсветке глаз ближним инфракрасным излучением [2]. С их помощью

фиксируется положение центра зрачка, а также отражения света на роговице глаза, так называемые блики. Анализ относительного смещения этих точек позволяет определить направление взгляда с высокой точностью.

Использование инфракрасного диапазона обеспечивает ряд преимуществ. Такой свет невидим для человеческого глаза, не вызывает дискомфорта и сохраняет естественное поведение человека. Кроме того, работа в инфракрасном диапазоне отличается высокой устойчивостью к изменениям внешнего освещения, что позволяет получать стабильное качество измерений.

Устройства отслеживания взгляда можно разделить на несколько типов в зависимости от задач и условий использования. Наиболее распространёнными являются экранные айтрекеры, которые подключаются к компьютеру и позволяют фиксировать направление взгляда пользователя при взаимодействии с информацией на экране. Другим распространённым типом являются носимые айтрекеры, выполненные в виде очков, такие устройства обеспечивают большую свободу движений и используются для анализа поведения человека при взаимодействии с окружающей средой. Данные устройства показаны на рисунках 1 и 2.



Рисунок 1. Экранный айтрекер



Рисунок 2. Носимый айтрекер

Программный подход не требует использования специализированного оборудования, благодаря чему снижаются стоимость внедрения и сложность интеграции технологии в существующие цифровые продукты. Кроме того, использование стандартных камер позволяет проводить исследования в дистанционном формате. Вместе с тем точность программных методов, как правило, уступает точности аппаратных систем и в значительной степени зависит от качества камеры, освещения и положения пользователя относительно устройства. По этой причине программные решения наиболее широко применяются в маркетинговых исследованиях, UX-аналитике и других задачах, не требующих лабораторной точности измерений. Сравнение данных подходов приведено в таблице 1.

Таблица 1. Сравнение подходов к отслеживанию взгляда

Критерий	Аппаратный подход	Программный подход
Оборудование	Специализированные айтрекеры	Обычные камеры
Принцип работы	Анализ смещения зрачка относительно бликов на роговице глаза	Алгоритмы компьютерного зрения
Стоимость внедрения	Высокая	Низкая

Точность	Высокая, менее 1°	Средняя, около 5°-7° (сильно зависит от условий проведения исследования)
Область применения	Научные и медицинские исследования	Маркетинг, UX-аналитика

Отслеживание взгляда с помощью нейросетей

Развитие методов глубокого обучения существенно расширило возможности программного отслеживания взгляда. Прогресс в данной области связан с появлением крупных обучающих наборов данных, увеличением доступных вычислительных ресурсов и совершенствованием нейросетевых архитектур. Это позволило повысить точность моделей и их устойчивость к изменениям освещения, положения головы и параметров используемых камер.

В основе большинства современных методов оценки направления взгляда лежат свёрточные нейронные сети. Они позволяют автоматически извлекать признаки из изображений глаз и лица и устанавливать зависимость между полученным визуальным представлением и направлением взгляда. Далее рассмотрены архитектуры iTracker, Special Weights CNN и L2CS-Net, отражающие основные этапы развития нейросетевых методов отслеживания взгляда.

Модель iTracker, представленная в 2016 году, стала одной из первых архитектур, продемонстрировавших возможность применения свёрточных нейронных сетей для отслеживания взгляда на мобильных устройствах [3]. Для её обучения был сформирован набор данных GazeCapture, включающий более 2,5 млн изображений, полученных при участии около 1,5 тыс. человек. Сбор данных проводился на различных устройствах, при изменяющемся освещении и свободном положении головы, что позволило повысить обобщающую способность модели.

Основная идея iTracker заключается в использовании не только изображений глаз, но и информации о положении лица относительно камеры. Архитектура данной модели показана на рисунке 3.

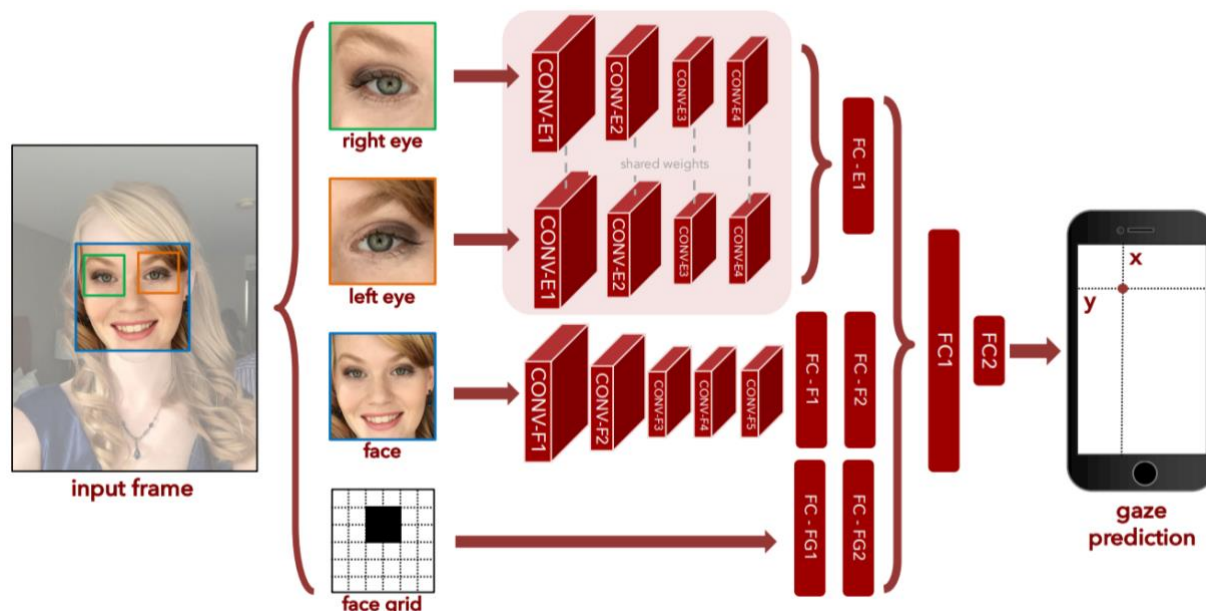


Рисунок 3. Архитектура iTracker

Архитектура имеет четыре входа: изображения левого и правого глаза, изображение лица и маску, характеризующую положение лица в кадре. Изображения глаз и лица обрабатываются отдельными свёрточными ветвями, предназначенными для извлечения признаков соответствующих областей. Затем полученные представления объединяются с информацией о положении лица и поступают в полносвязный регрессионный блок. Выходом модели являются координаты точки взгляда на экране устройства.

В 2017 году была предложена архитектура Special Weights CNN, позволяющая оценивать направление взгляда по изображению лица целиком без предварительного выделения областей глаз [4]. В зависимости от постановки задачи модель может предсказывать двумерные координаты точки взгляда либо нормированный трёхмерный вектор его направления.

Представление взгляда в виде трёхмерного вектора не связано с параметрами конкретного экрана и позволяет определять точку пересечения

направления взгляда с заданной плоскостью. Однако для решения такой задачи требуется предварительная нормализация входного изображения.

Нормализация предназначена для приведения изображений к единому пространству с фиксированными параметрами виртуальной камеры и одинаковым положением опорной точки. Для этого выполняются поворот изображения, изменение масштаба и преобразование расстояния до выбранной точки. Такая обработка снижает вариативность входных данных, связанную с положением пользователя и параметрами съёмки, а также обеспечивает согласованность обучающей разметки.

Ключевым элементом архитектуры является блок специальных пространственных весов, определяющий значимость отдельных областей лица. Он усиливает признаки, наиболее информативные для оценки взгляда, и подавляет вклад менее значимых участков изображения. Таким образом, модель самостоятельно выделяет области лица, оказывающие наибольшее влияние на результат. Архитектура данной модели показана на рисунке 4.

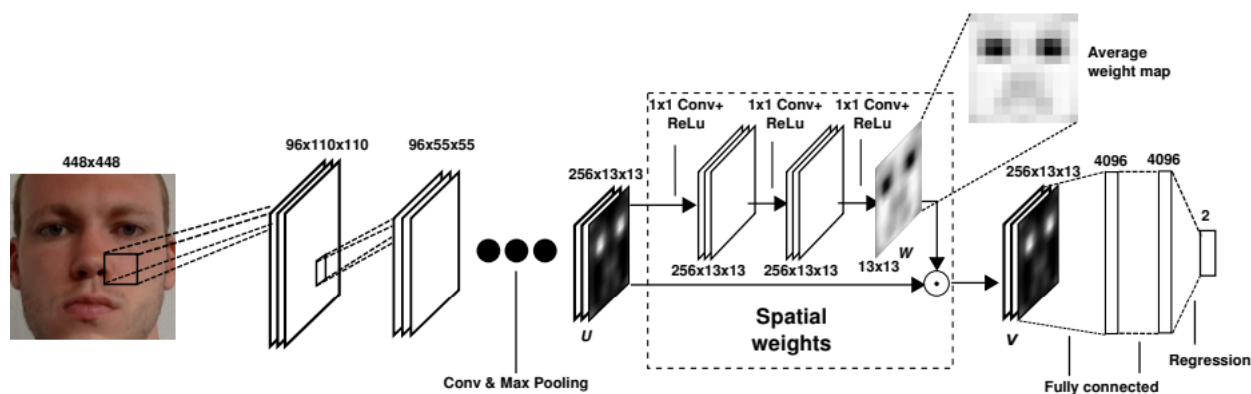


Рисунок 4. Архитектура Special Weights CNN

В качестве основы модели используется свёрточная сеть AlexNet, предназначенная для извлечения первичных признаков из изображения. Полученное признаковое представление поступает в блок пространственных весов, который с помощью последовательности свёрточных операций формирует карту значимости. Данная карта поэлементно умножается на исходное представление, усиливая наиболее

информативные области. На завершающем этапе скорректированные признаки передаются в полносвязный регрессионный блок, формирующий двумерные координаты точки фиксации или трёхмерный вектор направления взгляда.

В 2022 году была представлена модель L2CS-Net [5]. Её ключевой особенностью является раздельное предсказание горизонтального и вертикального углов направления взгляда. Для каждого угла используется самостоятельная выходная ветвь, что позволяет учитывать различия в распределении ошибок и независимо оптимизировать соответствующие предсказания.

В модели применяется подход, объединяющий классификацию и регрессию. Сначала диапазон значений каждого угла разделяется на равномерные интервалы, после чего сеть формирует вероятностное распределение принадлежности значения к соответствующим интервалам. Затем на основе полученного распределения восстанавливается непрерывное значение угла. Такое сочетание позволяет предварительно локализовать искомое значение и снизить чувствительность модели к отдельным выбросам. Архитектура данной модели показана на рисунке 5.

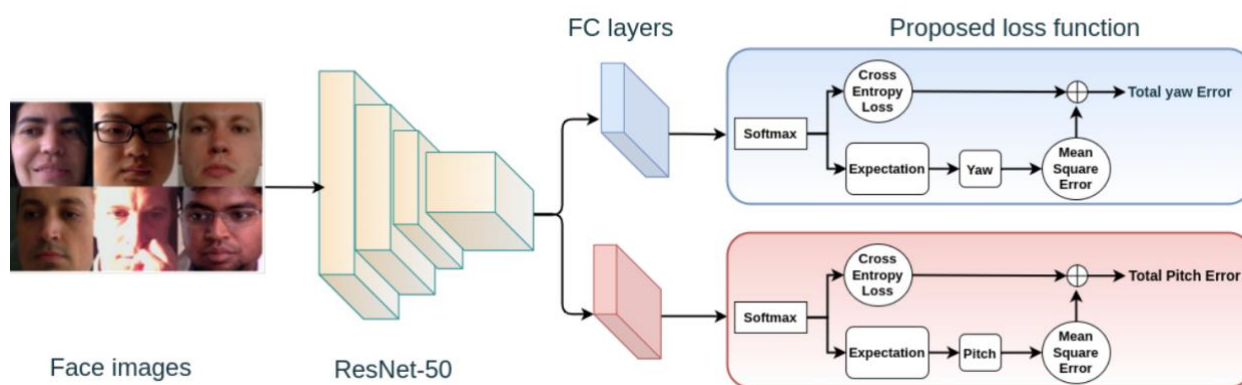


Рисунок 5. Архитектура L2CS-Net

В качестве основы архитектуры используется предварительно обученная свёрточная сеть ResNet-50, извлекающая признаки из изображения лица [6]. После общего блока извлечения признаков

располагаются две независимые выходные ветви, предназначенные для предсказания горизонтального и вертикального углов. Каждая ветвь реализует сочетание классификации и регрессии и формирует собственную составляющую функции потерь. Общая функция потерь определяется как сумма потерь двух ветвей и оптимизируется совместно в процессе обучения.

Для сопоставления рассмотренных архитектур используются средняя угловая и средняя позиционная ошибки. Сравнение архитектур приведено в таблице 4.

Таблица 4. Сравнение архитектур нейросетей

Датасет	Характеристики датасета	Метрика качества	iTracker	Special Weights CNN	L2CS-Net
GazeCapture	2,5 млн изображений, 1,5 тыс. участников	Средняя позиционная ошибка (см)	1,34 (телефон), 2,12 (планшет)	—	—
MPIIGaze	210 тыс. изображений, 15 участников	Средняя позиционная ошибка (см)	4,57	4,20	—
		Средняя угловая ошибка (°)	6,2	4,8	—
EYEDIAP	90 видеозаписей, 16 участников	Средняя позиционная ошибка (см)	13,0	8,56	—
		Средняя угловая ошибка (°)	8,3	6,0	—
MPIIFaceGaze	38 тыс. изображений	Средняя угловая ошибка (°)	—	—	3,92

Выводы

Сравнение подходов показало, что аппаратные системы обеспечивают более высокую точность и стабильность измерений, однако требуют

специализированного оборудования и значительных затрат. Программные методы уступают им по точности, но отличаются доступностью, простотой интеграции и возможностью использования стандартных камер в дистанционных исследованиях.

Анализ архитектур iTracker, Special Weights CNN и L2CS-Net показал, что их результаты зависят от используемых наборов данных, условий съёмки и выбранных метрик качества. Поэтому прямое сравнение моделей затруднено, а основным ограничением программного отслеживания взгляда остаётся снижение точности при переносе нейросетей из контролируемых условий в реальные сценарии применения.

Список литературы:

1. Изучение способов использования компьютерного зрения для определения направления взгляда [Электронный ресурс]. URL: <https://www.ultralytics.com/ru/blog/exploring-how-computer-vision-can-be-used-for-gaze-detection> (дата обращения: 22.09.2025).

2. How do eye trackers work [Электронный ресурс]. URL: <https://www.tobii.com/resource-center/learn-articles/how-do-eye-trackers-work> (дата обращения: 22.09.2025).

3. Krafska K., Khosla A., Kellnhofer P. et al. Eye Tracking for Everyone [Электронный ресурс] – URL: <https://arxiv.org/pdf/1606.05814> (дата обращения: 27.10.2025).

4. Zhang X., Sugano Y., Fritz M., Bulling A. Full-Face Appearance-Based Gaze Estimation [Электронный ресурс]. URL: https://www.collaborative-ai.org/publications/zhang17_cvprw.pdf (дата обращения: 27.10.2025).

5. Hsu G.S., Wang Y.C., Wang C.S. L2CS-Net: Fine-Grained Gaze Estimation in Unconstrained Environments [Электронный ресурс]. URL: <https://arxiv.org/pdf/2203.03339> (дата обращения: 27.10.2025).

6. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition [Электронный ресурс]. URL: <https://arxiv.org/pdf/1512.03385> (дата обращения: 27.10.2025).