Анализ статей (<https://paperswithcode.com/task/gaze-estimation>)

**Анализ датасетов и лучших моделей:**

Описаны основные требования к исходным изображениям, показан принцип работы софта для получения изображений

# [**Fine-Grained Head Pose Estimation Without Keypoints**](https://paperswithcode.com/paper/fine-grained-head-pose-estimation-without)**:**

Описаны данные о положении головы. Представлен способ определения позы путем обучения сверточной нейронной сети с множественными потерями на 300W-LP, большом синтетически расширенном наборе данных, для прогнозирования внутренних углов Эйлера (рыскание, тангаж и крен) непосредственно на основе интенсивности изображения посредством объединенной позы.

# [**Learning from Simulated and Unsupervised Images through Adversarial Training**](https://paperswithcode.com/paper/learning-from-simulated-and-unsupervised)**:**

Предложено обучение с имитацией + неконтролируемым обучением (S+U), где задача состоит в том, чтобы изучить модель для повышения реалистичности выходных данных симулятора с использованием немаркированных реальных данных, сохраняя при этом информацию аннотаций из симулятора. метод обучения S+U, который использует состязательную сеть, аналогичную генеративно-состязательным сетям (GAN), но с синтетическими изображениями в качестве входных данных вместо случайных векторов.

# [**Appearance-based Gaze Estimation With Deep Learning: A Review and Benchmark**](https://paperswithcode.com/paper/appearance-based-gaze-estimation-with-deep)**:**

В этой статье представлен всесторонний обзор методов оценки взгляда на основе внешнего вида с глубоким обучением.

# 

# [**Appearance-Based Gaze Estimation in the Wild**](https://paperswithcode.com/paper/appearance-based-gaze-estimation-in-the-wild):

В этой работе изучена оценка взгляда на основе внешнего вида. представлен набор данных MPIIGaze с подробным описанием сбора этих данных

# [**MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation**](https://paperswithcode.com/paper/mpiigaze-real-world-dataset-and-deep)**:**

Работа об устранении ограничений MPIIGaze.

# [**It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation**](https://paperswithcode.com/paper/its-written-all-over-your-face-full-face)**:**

предложен метод, основанный на внешнем виде, который, в отличие от давнего направления работы в компьютерном зрении, принимает в качестве входных данных только изображение анфас. метод анфас значительно превосходит современные методы оценки взгляда как в 2D, так и в 3D, достигая улучшений до 14,3% в MPIIGaze и 27,7% в EYEDIAP для оценки взгляда в 3D, не зависящей от человека

**3DGazeNet: Generalizing Gaze Estimation with Weak-Supervision from Synthetic Views:**

В этой работе предложено обучить общие модели оценки взгляда, которые могут быть непосредственно использованы в новых средах без адаптации. использовано наблюдение о том, что оценка положения головы, тела и рук выигрывает от пересмотра их как точного прогнозирования координат 3D, и аналогичным образом выражаем оценку взгляда как регрессию плотных сеток 3D для глаз.

[**Eye Tracking for Everyone**](https://paperswithcode.com/paper/eye-tracking-for-everyone)**:**

Описано улучшенное программное обеспечение для отслеживания глаз, которое работает на обычном оборудовании, таком как мобильные телефоны и планшеты, без необходимости в дополнительных датчиках или устройствах.

# [**Faster gaze prediction with dense networks and Fisher pruning**](https://paperswithcode.com/paper/faster-gaze-prediction-with-dense-networks)**:**

Исследуется предсказание фиксации человека на изображениях за счет использования глубоких представлений, которые были предварительно обучены распознаванию объектов.

**Recurrent CNN for 3D Gaze Estimation using Appearance and Shape Cues:**

В этой статье решаем проблему независимой от позы лица и головы оценки 3D взгляда с удаленных камер с использованием мультимодальной рекуррентной сверточной нейронной сети (CNN).

**Learning to Find Eye Region Landmarks for Remote Gaze Estimation in Unconstrained Settings:**

представлен новый метод, основанный на обучении, для определения положения ориентиров в области глаз, который позволяет традиционным методам конкурировать с новейшими методами, использующими внешний вид.

**RITnet: Real-time Semantic Segmentation of the Eye for Gaze Tracking:**

Решение проблем сегментации глаз. представлена модель RITnet, которая представляет собой глубокую нейронную сеть, сочетающую в себе U-Net и DenseNet. Размер RITnet составляет менее 1 МБ, а точность семантической сегментации OpenEDS 2019 года составляет 95,3%. Используя GeForce GTX 1080 Ti, RITnet отслеживает более 300 Гц, что позволяет использовать приложения для отслеживания взгляда в реальном времени.

# [**Towards High Performance Low Complexity Calibration in Appearance Based Gaze Estimation**](https://paperswithcode.com/paper/geddnet-a-network-for-gaze-estimation-with)**:**

В данной работе расширен предыдущий метод разложения взгляда, который разделяет угол взгляда на две составляющие: независимую от субъекта оценку взгляда по изображению и зависящее от субъекта смещение. Мы исследуем влияние количества целей взгляда, количества использованных изображений на каждую цель и количества положений головы в калибровочных данных с использованием нового набора данных NISLGaze.

# [**Self-Learning Transformations for Improving Gaze and Head Redirection**](https://paperswithcode.com/paper/self-learning-transformations-for-improving):

В данном исследовании предложена новая генеративная модель изображений лиц, способную создавать высококачественные изображения с подробным контролем над взглядом глаз и углами ориентации головы.

# [**TinyTracker: Ultra-Fast and Ultra-Low-Power Edge Vision In-Sensor for Gaze Estimation**](https://paperswithcode.com/paper/tinytracker-ultra-fast-and-ultra-low-power)**:**

В данном исследовании оценивавется использование платформы машинного зрения "ИИ в сенсоре" IMX500 от Sony для создания ультрабыстрых и энергоэффективных приложений периферийного зрения. сравнивается IMX500 с другими платформами, такими как Google Coral Dev Micro и Sony Spresense, на примере оценки взгляда. Представлен TinyTracker - высокоэффективную модель для оценки взгляда, которая обеспечивает значительное уменьшение размера без потери точности (максимум 0,16 см при полном квантовании).

**Ecological Sampling of Gaze Shifts:**

Предложенная здесь стратегия выборки позволяет имитировать фундаментальное свойство механизма наведения глаз: *выбор следующего взгляда в любой момент времени не является полностью детерминированным, но и не является полностью случайным.*

# [**Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction**](https://paperswithcode.com/paper/pupil-an-open-source-platform-for-pervasive)**:**

*представлена Pupil — доступная, недорогая и расширяемую платформу с открытым исходным кодом для мобильного отслеживания глаз и взаимодействия на основе взгляда*. Программное обеспечение и графический интерфейс не зависят от платформы и включают в себя самые современные алгоритмы для обнаружения и отслеживания зрачков в реальном времени, калибровки и точной оценки взгляда.

# [**Deep Pictorial Gaze Estimation**](https://paperswithcode.com/paper/deep-pictorial-gaze-estimation)**:**

представлена новая архитектура глубокой нейронной сети, специально разработанная для задачи оценки взгляда по входным данным одного глаза. Наш подход обеспечивает более высокую точность, чем современный, и устойчив к изменениям взгляда, положения головы и качества изображения.

# [**RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments**](https://paperswithcode.com/paper/rt-gene-real-time-eye-gaze-estimation-in)**:**

представлен новый алгоритм реального времени, включающий глубокие сверточные нейронные сети на основе внешнего вида с повышенной способностью справляться с разнообразными изображениями в новом наборе данных.

# [**Learning to Zoom: a Saliency-Based Sampling Layer for Neural Networks**](https://paperswithcode.com/paper/learning-to-zoom-a-saliency-based-sampling)**:**

Мы вводим слой искажения на основе значимости для сверточных нейронных сетей, который помогает улучшить пространственную выборку входных данных для данной задачи. *Эффект слоя заключается в эффективной оценке того, как производить выборку из исходных данных, чтобы повысить производительность задачи*.

**Hybrid coarse-fine classification for head pose estimation:**

В этой статье, чтобы выполнить оценку без ориентиров лица, объединены результаты грубой и точной регрессии для создания глубокой сети. Используя больше единиц квантования для углов, точный классификатор обучается с помощью других вспомогательных грубых единиц. Интегрирующая регрессия используется для получения окончательного прогноза. Предлагаемый подход оценивается по трем сложным критериям. Он соответствует последнему слову техники на AFLW2000, BIWI и хорошо работает на AFLW. Код был опубликован на Github.

# **Photo-Realistic Monocular Gaze Redirection Using Generative Adversarial Networks:**

Перенаправление взгляда. Существующим методам не хватает способности генерировать воспринимаемые правдоподобные изображения. В этой работе мы представляем новый метод решения этой проблемы за счет использования генеративно-состязательного обучения для синтеза изображения глаза, обусловленного целевым направлением взгляда.

# 

# **Unsupervised Learning of Eye Gaze Representation from the Web:**

В этой статье мы предлагаем метод, основанный на обучении без учителя, для оценки области взгляда.

**Few-Shot Adaptive Gaze Estimation:**

предложена новая основа для адаптивной оценки GaZE с несколькими кадрами (FAZE) для изучения сетей взглядов, специфичных для человека, с очень небольшим количеством (менее или равным 9) калибровочных образцов.

**Mixed Effects Neural Networks (MeNets) With Applications to Gaze Estimation:**

Цель этой статьи — адаптировать эти идеи «смешанных эффектов» из статистики в архитектуре глубокой нейронной сети для оценки взгляда на основе изображений глаз.

# **Gaze360: Physically Unconstrained Gaze Estimation in the Wild**

В этой работе представляем Gaze360, крупномасштабный набор данных для отслеживания взгляда и метод надежной трехмерной оценки взгляда на изображениях без ограничений. Предлагаемая трехмерная модель взгляда расширяет существующие модели, включая временную информацию и прямой вывод оценки неопределенности взгляда.

# **RT-BENE: A Dataset and Baselines for Real-Time Blink Estimation in Natural Environments**

*многие методы оценки взгляда обычно предполагают, что глаза субъекта открыты; для закрытых глаз эти методы дают нерегулярные оценки взгляда.* Здесь рассматривается это предположение. Далее представлены базовые методы, которые позволяют *обнаруживать моргание* с помощью сверточных нейронных сетей.

# **Gaze Preserving CycleGANs for Eyeglass Removal & Persistent Gaze Estimation**

*Оценка направления взгляда — наиболее очевидный способ оценить состояние водителя в идеальных условиях*. В этом исследовании предложено решение проблем, возникающих в реальном мире(освещение и тп.).

# **ETH-XGaze: A Large Scale Dataset for Gaze Estimation under Extreme Head Pose and Gaze Variation**

В этой статье мы предлагаем *новый набор данных для оценки взгляда* под названием ETH-XGaze, состоящий из более чем миллиона изображений с высоким разрешением различных взглядов при экстремальных положениях головы.

# **360-Degree Gaze Estimation in the Wild Using Multiple Zoom Scales**

В этой работе мы разрабатываем модель, которая *имитирует способность человека оценивать взгляд путем агрегирования сфокусированных взглядов*, каждый из которых имеет разный уровень увеличения области лица. Модель позволяет избежать необходимости извлекать четкие пятна на глазах и в то же время решает еще одну важную проблему изменения масштаба лица для оценки взгляда в дикой природе.

# Studying Person-Specific Pointing and Gaze Behavior for Multimodal Referencing of Outside Objects from a Moving Vehicle

Указание руки и взгляд широко исследовались в автомобильных приложениях для выбора объектов и привязки к ним. Статистический анализ показывает значительные различия в индивидуальном поведении в зависимости от местоположения объекта (т. е. с правой стороны водителя по сравнению с левой), окружения объекта, режима вождения (т. е. автономного и обычного вождения), а также от направления и продолжительности взгляда, что закладывает основу для адаптивный подход к пользователю

# Cross-Encoder for Unsupervised Gaze Representation Learning

Чтобы обучить 3D-оценщики взгляда без слишком большого количества аннотаций, мы предлагаем систему обучения без учителя Cross-Encoder. Она распутывает черты, используя механизм замены скрытого кода на парах изображений, согласованных с глазами, и парах изображений, похожих на взгляд. Cross-Encoder обучен реконструировать каждое изображение в паре, совместимой с глазами, в соответствии с особенностью его взгляда и особенностью глаза другого, но реконструировать каждое изображение в паре, похожей на взгляд, в соответствии с особенностью его глаза и особенностью взгляда другого.

# Goal-Oriented Gaze Estimation for Zero-Shot Learning

Обучение с нулевым выстрелом (ZSL) направлено на распознавание новых классов путем передачи семантических знаний от видимых классов к невидимым классам. Распознавая невидимые изображения, человек также автоматически смотрит на области с определенной семантической подсказкой. Мы стремимся предсказать фактическое местоположение взгляда человека, чтобы получить области зрительного внимания для распознавания нового объекта, руководствуясь описанием атрибута.

# PureGaze: Purifying Gaze Feature for Generalizable Gaze Estimation

Методы оценки взгляда изучают взгляд глаз по чертам лица. Мы устраняем факторы, не имеющие отношения к взгляду, такие как освещенность и идентичность, чтобы улучшить междоменную производительность. Мы разрабатываем самозащитную систему «подключи и работай» для очистки функций взгляда. Эта структура напрямую и значительно расширяет не только наши базовые показатели, но и существующие методы оценки взгляда.

# Weakly-Supervised Physically Unconstrained Gaze Estimation

Мы решаем ранее не исследованную проблему слабо контролируемой оценки взгляда по видеозаписям человеческих взаимодействий. Мы опираемся на понимание того, что сильные геометрические ограничения, связанные с взглядом, существуют, когда люди выполняют действие «смотря друг на друга».

# GOO: A Dataset for Gaze Object Prediction in Retail Environments

Одно из наиболее фундаментальных и насыщенных информацией действий человека — это смотреть на объекты. Существующие наборы данных, связанные с взглядом, аннотируют только просматриваемый пиксель, а не границы конкретного интересующего объекта. GOO состоит из большого набора синтетических изображений людей, рассматривающих объекты в розничной торговле.

Gaze Estimation using Transformer

Недавние работы доказали эффективность трансформаторов во многих задачах компьютерного зрения. Однако эффективность преобразователей при оценке взгляда до сих пор не изучена. В этой статье мы используем трансформаторы и оцениваем их эффективность для оценки взгляда. Эксперименты показывают, что гибридный трансформатор может достичь высочайшей производительности во всех тестах при предварительном обучении.

# The Story in Your Eyes: An Individual-difference-aware Model for Cross-person Gaze Estimation

Мы предлагаем новый метод уточнения задачи прогнозирования взгляда перекрестного человека с использованием изображений глаз/лица только путем явного моделирования специфичных для человека различий. Прогнозируя надежность текущих изображений глаз/лица, наша виртуальная машина может идентифицировать недействительные образцы, например изображения моргания глаз, и уменьшать их влияние в процессе моделирования. Затем наш модуль SC и PT учится компенсировать различия только на действительных образцах.

# Gaze Estimation with an Ensemble of Four Architectures

В этой статье представлен метод оценки взгляда по изображениям лица. Мы обучаем несколько оценщиков взгляда, используя четыре различные сетевые архитектуры. Затем мы выбираем шесть лучших оценщиков и объединяем их прогнозы с помощью линейной комбинации.

# Generalizing Gaze Estimation with Outlier-guided Collaborative Adaptation

Глубокие нейронные сети значительно улучшили точность оценки взгляда на основе внешнего вида. Однако она по-прежнему страдает от неудовлетворительной производительности при обобщении обученной модели на новые области, например, невидимую среду или людей. В этой статье мы предлагаем систему адаптации взгляда «подключи и работай» (PnP-GA), которая представляет собой ансамбль сетей, которые обучаются совместно под руководством посторонних факторов.

# Automatic Gaze Analysis: A Survey of Deep Learning based Approaches

Есть несколько открытых вопросов, в том числе, каковы важные сигналы для интерпретации направления взгляда в неограниченной среде без предварительных знаний и как их кодировать в режиме реального времени. Наш анализ показывает, что разработка надежного и универсального метода анализа взгляда по-прежнему требует решения реальных проблем, таких как неограниченная настройка и обучение с меньшим контролем. В заключение мы обсудим будущие направления исследований по разработке реальной системы анализа взгляда

# High-Accuracy Gaze Estimation for Interpolation-Based Eye-Tracking Methods

В этом исследовании изучается влияние местоположения глазной камеры, связанное с точностью и точностью методов отслеживания глаз на основе интерполяции. Наши эксперименты показывают, что расположение камеры-глаза в сочетании с некомпланарностью плоскости глаза искажает распределение характеристик глаза, когда камера-глаз находится далеко от оптической оси глаза. В этой статье предлагаются методы геометрического преобразования для изменения распределения характеристик глаза на основе виртуального выравнивания глаза-камеры по центру оптической оси глаза.

# FLAME: Facial Landmark Heatmap Activated Multimodal Gaze Estimation

Оценка 3D-взгляда заключается в прогнозировании направления взгляда человека в 3D-пространстве. Независимым от человека моделям не хватает точности из-за анатомических различий субъектов, тогда как калиброванные методы, специфичные для человека, накладывают строгие ограничения на масштабируемость.

# GaTector: A Unified Framework for Gaze Object Prediction

Предсказание объектов взгляда — это недавно предложенная задача, целью которой является обнаружение объектов, на которые смотрят люди. Он имеет большое прикладное значение, но до сих пор не имеет единой структуры решения. Интуитивное решение — включить ветвь обнаружения объектов в существующий метод прогнозирования взгляда. В этой статье мы создаем новую среду под названием GaTector для унифицированного решения проблемы прогнозирования объектов взгляда.

# Self-Paced Deep Regression Forests with Consideration of Ranking Fairness

Глубокие дискриминационные модели (DDM), например, глубокие регрессионные леса и глубокие леса решений, в последнее время широко изучаются для решения таких проблем, как оценка возраста лица, оценка положения головы и т. д. Из-за нехватки хорошо размеченных данных, которые не имеют проблемы шума и несбалансированного распределения, изучение DDM всегда является сложной задачей. Обширные экспериментальные результаты по трем задачам компьютерного зрения, а именно: оценка возраста лица, оценка положения головы и оценка взгляда, показывают, что наш новый метод значительно улучшает производительность как с точки зрения точности, так и объективности

# Dynamic 3D Gaze From Afar: Deep Gaze Estimation From Temporal Eye-Head-Body Coordination

Мы представляем новый метод и набор данных для трехмерной оценки взгляда свободно движущегося человека на расстоянии, обычно в режиме наблюдения. В таких случаях глаза не могут быть четко видны из-за окклюзии и недостаточного разрешения. Мы представляем новый обширный набор данных, состоящий из видео наблюдения с указанием трехмерных направлений взгляда, снятых в 5 сценах внутри и снаружи помещения.

# Resolving Camera Position for a Practical Application of Gaze Estimation on Edge Devices

Большинство исследований по оценке взгляда работают только при условии, что камера идеально фиксирует взгляд глаз. В них не указано, как правильно настроить камеру для данного положения человека. В этой статье мы проводим исследование оценки взгляда с логическим положением установки камеры. Мы также применяем наши исследования на практике, используя недорогие периферийные устройства с реалистичным сценарием. То есть сначала мы настраиваем торговую среду, в которой хотим отслеживать поведение покупателей при взгляде. Для этой настройки требуется оптимальное положение камеры, чтобы поддерживать точность оценки на основе существующих исследований оценки взгляда. Затем мы применяем современную методику оценки обучающего взгляда с нескольких кадров, чтобы сократить количество обучающих выборок на этапе вывода. В ходе эксперимента мы проводим наше реализованное исследование на NVIDIA Jetson TX2 и достигаем разумной скорости, 12 кадров в секунду, что выше по сравнению с нашей эталонной работой, без существенного снижения точности оценки взгляда.

# L2CS-Net: Fine-Grained Gaze Estimation in Unconstrained Environments

Оценка взгляда в дикой природе по-прежнему остается сложной проблемой из-за уникальности внешнего вида глаз, условий освещения, а также разнообразия поз головы и направлений взгляда. В этой статье мы предлагаем надежную модель на основе CNN для прогнозирования взгляда в неограниченных условиях.

# Towards Self-Supervised Gaze Estimation

Недавние методы самоконтроля на основе совместного внедрения превзошли стандартные контролируемые подходы к различным задачам распознавания изображений, таким как классификация изображений. В этой работе мы предлагаем SwAT, эквивариантную версию онлайн-подхода SwAV с самоконтролем на основе кластеризации, чтобы изучить более информативные представления для оценки взгляда.

# EllSeg-Gen, towards Domain Generalization for head-mounted eyetracking

Алгоритмы часто не могут идентифицировать такие особенности, как радужная оболочка и центр тяжести зрачка, при наличии отражающих артефактов и окклюзий. Мы сравниваем производительность одной модели, обученной на нескольких наборах данных, с пулом моделей, обученных на отдельных наборах данных.

Multistream Gaze Estimation with Anatomical Eye Region Isolation by Synthetic to Real Transfer Learning

Мы предлагаем новый нейронный конвейер MSGazeNet, который изучает представления взгляда, используя информацию об анатомии глаза через многопоточную платформу. Предлагаемое нами решение состоит из двух компонентов: первого - сети для выделения анатомических областей глаза и второго - сети для многопоточного оценивания взгляда. Выделение области глаза выполняется с помощью сети в стиле U-Net, которую мы обучаем с использованием синтетического набора данных, содержащего маски области глаза для видимого глазного яблока и области радужной оболочки. Синтетический набор данных, используемый на этом этапе, получен с помощью симулятора UnityEyes и состоит из 80 000 изображений глаз. После обучения сеть выделения области глаза затем переносится в реальную область для создания масок для изображений реальных глаз. Для успешного переноса мы используем рандомизацию предметной области в процессе обучения, которая позволяет синтетическим изображениям извлекать выгоду из большей дисперсии с помощью дополнений, напоминающих артефакты. Сгенерированные маски области глаза вместе с необработанными изображениями глаз затем используются вместе в качестве многопоточных входных данных для нашей сети оценки взгляда, которая состоит из широких остаточных блоков. Выходные данные этих кодеров объединяются в канальном измерении перед подачей в слои регрессии взгляда. Мы оцениваем нашу платформу на трех наборах данных оценки взгляда и достигаем высоких показателей. Наш метод превосходит самый современный на 7,57% и 1,85% для двух наборов данных и дает конкурентоспособные результаты для другого. Мы также изучаем надежность нашего метода в отношении шума в данных и демонстрируем, что наша модель менее чувствительна к зашумленным данным. Наконец, мы проводим множество экспериментов, включая исследования абляции, чтобы оценить вклад различных компонентов и выбор дизайна в наше решение.

# 'Labelling the Gaps': A Weakly Supervised Automatic Eye Gaze Estimation

За последние несколько лет возрос интерес к интерпретации направления взгляда в неограниченной среде с ограниченным наблюдением. Из-за проблем с обработкой данных и аннотациями тиражирование метода оценки взгляда на другие платформы, такие как unconstrained outdoor или AR / VR, может привести к значительному снижению производительности из-за недостаточной доступности точно аннотированных данных для обучения модели. В этой статье мы исследуем интересную, но сложную проблему метода оценки взгляда с ограниченным количеством помеченных данных. Предлагаемый метод извлекает знания из помеченного подмножества с визуальными характеристиками; включая внешний вид, специфичный для личности, согласованность траектории взгляда и особенности движения. Учитывая траекторию взгляда, метод использует информацию о метках только для начального и конечного кадров последовательности взглядов. Расширение предлагаемого метода дополнительно снижает требования к маркированным рамкам до только начального кадра с незначительным снижением качества создаваемой этикетки. Мы оцениваем предлагаемый метод на четырех контрольных наборах данных (CAVE, TabletGaze, MPII и Gaze360), а также на видеороликах YouTube, которые просматриваются в Интернете. Предлагаемый нами метод сокращает затраты на аннотацию до 2,67% с минимальным влиянием на производительность; это указывает на потенциал нашей модели, позволяющий проводить оценку взгляда "в естественных условиях".

# Multimodal Across Domains Gaze Target Detection

В этой статье рассматривается проблема обнаружения цели взгляда на отдельных изображениях, снятых с точки зрения третьего лица. Мы представляем мультимодальную глубокую архитектуру, позволяющую определить, куда смотрит человек в сцене. Эта пространственная модель обучается на изображениях головы интересующего человека, картах сцены и глубины, представляющих обширную контекстную информацию. Наша модель, в отличие от нескольких известных моделей, не требует контроля углов взгляда, не полагается на информацию об ориентации головы и / или расположении глаз интересующего человека. Обширные эксперименты демонстрируют более высокую производительность нашего метода на нескольких контрольных наборах данных. Мы также исследовали несколько вариантов нашего метода, изменив совместное изучение мультимодальных данных. Некоторые варианты также превосходят некоторые известные образцы техники. Впервые в этой статье мы проверяем адаптацию домена для обнаружения целей gaze и расширяем возможности нашей мультимодальной сети для эффективного устранения разрыва в домене между наборами данных.

# LatentGaze: Cross-Domain Gaze Estimation through Gaze-Aware Analytic Latent Code Manipulation

Хотя в недавних методах оценки взгляда большое внимание уделяется тщательному выделению релевантных для взгляда черт из изображений лица или глаз, способы определения признаков, которые включают компоненты, релевантные для взгляда, были неоднозначными. Эта неясность заставляет модель изучать не только функции, относящиеся к взгляду, но и нерелевантные. В частности, это фатально для производительности при работе с несколькими наборами данных. Чтобы преодолеть эту сложную проблему, мы предлагаем метод аналитических манипуляций с учетом пристального взгляда, основанный на подходе, основанном на данных, с характеристиками распутывания generative adversarial network inversion, позволяющий выборочно использовать функции, относящиеся к пристальному взгляду, в латентном коде. Кроме того, используя процесс кодирования-генератора на основе GAN, мы перемещаем входное изображение из целевого домена в изображение исходного домена, о котором специалист по оценке пристального взгляда достаточно осведомлен. Кроме того, мы предлагаем исключить искажение взгляда в кодере, что предотвращает искажение информации о взгляде. Результаты эксперимента демонстрируют, что наш метод обеспечивает высочайшую точность оценки пристального взгляда в междоменных задачах оценки пристального взгляда.

# HAZE-Net: High-Frequency Attentive Super-Resolved Gaze Estimation in Low-Resolution Face Images

Хотя методы оценки взгляда были разработаны с использованием методов глубокого обучения, не было такого подхода, который был бы направлен на достижение точной производительности на изображениях лиц низкого разрешения с шириной пикселя 50 пикселей или менее. Чтобы устранить ограничение в сложных условиях низкого разрешения, мы предлагаем высокочастотную сеть оценки внимательного взгляда со сверхразрешением, т. е. HAZE-Net. Наша сеть улучшает разрешение входного изображения и улучшает особенности глаз и их границы с помощью предлагаемого модуля сверхразрешения, основанного на высокочастотном блоке внимания. Кроме того, наш модуль оценки взгляда использует высокочастотные компоненты глаза, а также глобальную карту внешнего вида. Мы также используем информацию о структурном расположении лиц для приблизительного определения положения головы. Результаты эксперимента показывают, что предлагаемый метод демонстрирует надежную оценку взгляда даже на изображениях лиц с низким разрешением 28x28 пикселей.

# Contrastive Representation Learning for Gaze Estimation

Самоконтролируемое обучение (SSL) стало распространенным методом изучения представлений в компьютерном зрении. Примечательно, что SSL использует контрастивное обучение, чтобы сделать визуальные представления инвариантными при различных преобразованиях изображений. Задача оценки взгляда, с другой стороны, требует не только инвариантности к различным внешним проявлениям, но и эквивалентности геометрическим преобразованиям. В этой работе мы предлагаем простую платформу обучения контрастивному представлению для оценки взгляда, названную GazeCLR (Gaze Contrastive Learning). GazeCLR использует данные из нескольких представлений для обеспечения эквивариантности и полагается на выбранные методы увеличения данных, которые не изменяют направления взгляда для изучения инвариантности. Наши эксперименты демонстрируют эффективность GazeCLR для нескольких настроек задачи оценки взгляда. В частности, наши результаты показывают, что GazeCLR улучшает производительность междоменной оценки взгляда и дает относительное улучшение на 17,2%. Более того, фреймворк GazeCLR конкурирует с самыми современными методами обучения представлению для оценки с нескольких попыток.

# One Eye is All You Need: Lightweight Ensembles for Gaze Estimation with Single Encoders

Точность оценки взгляда в последние годы быстро возросла. Однако в этих моделях часто не используются преимущества различных алгоритмов и методов компьютерного зрения (CV) (таких как небольшие сети ResNet и Inception networks, а также ансамблевые модели), которые, как было показано, улучшают результаты при решении других задач CV. Кроме того, большинство современных моделей оценки взгляда требуют использования либо обоих глаз, либо всего лица, в то время как в реальных данных не всегда оба глаза отображаются в высоком разрешении. Таким образом, мы предлагаем модель оценки взгляда, которая реализует архитектуры ResNet и Inception model и делает прогнозы, используя только изображение одного глаза. Кроме того, мы предлагаем сеть калибровки ансамбля, которая использует прогнозы из нескольких отдельных архитектур для прогнозирования конкретных объектов. Используя легкие архитектуры, мы достигаем высокой производительности в наборе данных GazeCapture при очень низком количестве параметров модели. При использовании двух глаз в качестве входных данных мы достигаем ошибки прогнозирования в 1,591 см на тестовом наборе без калибровки и 1,439 см с моделью калибровки ансамбля. Используя только один глаз в качестве входных данных, мы по-прежнему достигаем средней ошибки прогнозирования в 2,312 см на тестовом наборе без калибровки и 1,951 см с моделью калибровки ансамбля. Мы также заметили значительно меньшее количество ошибок на изображениях правого глаза в тестовом наборе, что может быть важно при разработке будущих инструментов, основанных на оценке взгляда.

# Source-Free Adaptive Gaze Estimation by Uncertainty Reduction

Оценка пристального взгляда в разных областях была недавно исследована, потому что обучающие данные обычно собираются в контролируемых условиях, в то время как обученные методы оценки пристального взгляда используются в реальных и разнообразных средах. Однако из-за соображений конфиденциальности и эффективности одновременный доступ к аннотированным исходным данным и прогнозируемым целевым данным может быть затруднен. В свете этого мы представляем неконтролируемый подход к адаптации домена без источника для оценки пристального взгляда, который адаптирует обученный источнику оценщик пристального взгляда к немаркированным целевым доменам без исходных данных. Мы предлагаем фреймворк для адаптации Gaze с уменьшением неопределенности (UnReGA), который обеспечивает адаптацию за счет уменьшения неопределенности как выборки, так и модели. Неопределенность выборки снижается за счет повышения качества изображений и упрощения их оценки взгляда, в то время как неопределенность модели снижается за счет минимизации дисперсии прогноза при одних и тех же входных данных. Обширные эксперименты проводятся над шестью междоменными задачами, демонстрирующими эффективность UnReGA и ее компонентов. Результаты показывают, что UnReGA превосходит другие современные междоменные методы оценки gaze по обоим протоколам, с исходными данными и без них

# Using Gaze for Behavioural Biometrics

Изложен принципиальный подход к анализу движений глаз для поведенческой биометрии. Подход основан на теории поиска пищи, которая обеспечивает надежную основу для выявления уникальности индивидуального поведения при движении глаз. Мы предлагаем комплексный процесс Орнштейна-Уленбека для количественной оценки сигнатуры разведки / эксплуатации, характеризующей поведение глаз-охотников. Показано, что соответствующие параметры составной модели, полученные на основе данных отслеживания взгляда с помощью байесовского анализа, дают подходящий набор признаков для биометрической идентификации; последнее в конечном итоге достигается с помощью классической техники классификации. Подтверждение концепции метода обеспечивается путем измерения эффективности его идентификации на общедоступном наборе данных. Доступны данные и код для воспроизведения анализов. В целом, мы утверждаем, что этот подход предлагает свежий взгляд как на анализ данных отслеживания движения глаз, так и на перспективные приложения в этой области.

# Towards Precision in Appearance-based Gaze Estimation in the Wild

Системы оценки внешнего вида по взгляду в последнее время продемонстрировали большой прогресс, однако эффективность этих методов зависит от наборов данных, используемых для обучения. Большинство существующих наборов данных для оценки взгляда, настроенных в интерактивных настройках, были зарегистрированы в лабораторных условиях, а те, которые были зарегистрированы в диких условиях, демонстрируют ограниченные изменения положения головы и освещения. Кроме того, мы заметили, что до сих пор мало внимания уделялось точным оценкам существующих подходов к оценке взгляда. В этой работе мы представляем большой набор данных для оценки взгляда, PARKS-Gaze, с более широким изменением положения головы и освещенности, а также с несколькими выборками для одной точки взгляда (PoG). Набор данных содержит 974 минуты данных от 28 участников с диапазоном положения головы 60 градусов как в направлении рыскания, так и по тангажу. Наши оценки внутри набора данных и между наборами данных, а также оценки точности показывают, что предлагаемый набор данных является более сложным и позволяет моделям обобщать данные о невидимых участниках лучше, чем существующие наборы данных в дикой природе.

# Precise localization of corneal reflections in eye images using deep learning trained on synthetic data

Мы представляем метод глубокого обучения для точной локализации центра единичного отражения от роговицы (CR) на изображении глаза. В отличие от предыдущих подходов, мы используем сверточную нейронную сеть (CNN), которая обучалась исключительно на имитированных данных. Преимущество использования только смоделированных данных заключается в том, что полностью исключается трудоемкий процесс ручного аннотирования, необходимый для контролируемого обучения на изображениях реальных глаз. Чтобы систематически оценивать точность нашего метода, мы сначала протестировали его на изображениях с имитацией отражений, размещенных на разном фоне и встроенных в различные уровни шума. Во-вторых, мы протестировали метод на высококачественных видео, снятых реальными глазами. Наш метод превзошел самые современные алгоритмические методы на изображениях реальных глаз со снижением пространственной точности на 35% и работает наравне с современными моделируемыми изображениями с точки зрения пространственной точности.Мы приходим к выводу, что наш метод обеспечивает точный метод локализации центра CR и обеспечивает решение проблемы доступности данных, которая является одним из важных распространенных препятствий при разработке моделей глубокого обучения для оценки взгляда. Благодаря превосходной локализации центра CR и простоте применения наш метод потенциально может повысить точность глазных трекеров на основе CR.

# Investigation of Architectures and Receptive Fields for Appearance-based Gaze Estimation

В связи с быстрым развитием технологии глубокого обучения в последнее десятилетие оценка взгляда на основе внешнего вида привлекла большое внимание как со стороны сообществ исследователей компьютерного зрения, так и взаимодействия человека с компьютером. Были предложены увлекательные методы с различными механизмами, включая мягкое внимание, жесткое внимание, асимметрию двух глаз, распутывание признаков, последовательность вращения и контрастивное обучение. Большинство этих методов используют в качестве входных данных одно лицо или несколько областей, однако базовая архитектура оценки взгляда изучена не полностью. В этой статье мы раскрываем тот факт, что настройка нескольких простых параметров архитектуры ResNet может превзойти большинство существующих современных методов для задачи оценки взгляда на трех популярных наборах данных. В результате наших обширных экспериментов мы пришли к выводу, что количество шагов, разрешение входного изображения и архитектура с несколькими регионами имеют решающее значение для производительности оценки взгляда, в то время как их эффективность зависит от качества входного изображения лица. Мы получаем самые современные показатели для трех наборов данных с 3,64 для ETH-XGaze, 4,50 для MPIIFaceGaze и 9,13 для ошибки оценки взгляда на 360 градусов, используя ResNet-50 в качестве основы.

# DVGaze: Dual-View Gaze Estimation

Методы оценки взгляда оценивают взгляд по внешнему виду лица с помощью одной камеры. Однако из-за ограниченного обзора с одной камеры запечатленный внешний вид лица не может предоставить полную информацию о лице и, таким образом, усложняет задачу оценки взгляда. В последнее время устройства с камерами быстро обновляются. Двойные камеры доступны для пользователей и были интегрированы во многие устройства. Эта разработка предполагает, что мы можем еще больше повысить эффективность оценки пристального взгляда с помощью оценки пристального взгляда в режиме двойного обзора. В этой статье мы предлагаем сеть оценки взгляда в двух ракурсах (DV-Gaze). DV-Gaze оценивает направления взгляда в двух ракурсах по паре изображений. Сначала мы предлагаем блок интерактивной свертки с двойным обзором (DIC) в DV-Gaze. Блоки DIC обмениваются информацией с двойным обзором во время свертки в нескольких масштабах объектов. Он объединяет объекты с двойным обзором по эпиполярным линиям и компенсирует исходную функцию с помощью объединенной функции. Далее мы предлагаем преобразователь с двойным обзором для оценки взгляда с помощью объектов с двойным обзором. Позы камеры закодированы для отображения информации о положении в трансформаторе. Мы также рассматриваем геометрическую связь между направлениями взгляда в режиме двойного обзора и предлагаем потерю согласованности взгляда в режиме двойного обзора для DV-Gaze. DV-Gaze обеспечивает самую современную производительность в наборах данных ETH-XGaze и EVE.

# Gaze Estimation on Spresense

Оценка взгляда - ценная технология с многочисленными приложениями в таких областях, как взаимодействие человека и компьютера, виртуальная реальность и медицина. В этом отчете представлена реализация системы оценки gaze с использованием платы микроконтроллера Sony Spresense и исследуются ее показатели по задержке, MAC / циклу и энергопотреблению. В отчете также содержится информация об архитектуре системы, включая используемую модель оценки gaze. Дополнительно представлена демонстрация системы, демонстрирующая ее функциональность и производительность. Наша облегченная модель TinyTrackerS имеет размер всего 169 КБ, использует параметры 85,8 кб и работает на платформе Spresense со скоростью 3 кадра в секунду.

# LEyes: A Lightweight Framework for Deep Learning-Based Eye Tracking using Synthetic Eye Images

Глубокое обучение укрепило методы оценки взгляда, но внедрению в реальном мире препятствовали неадекватные обучающие наборы данных. Эта проблема усугубляется как вызванными аппаратным обеспечением вариациями изображений глаз, так и врожденными биологическими различиями между записанными участниками, что приводит к различиям как на уровне объектов, так и на уровне пикселей, что препятствует обобщаемости моделей, обученных на конкретных наборах данных. Хотя синтетические наборы данных могут быть решением, их создание требует как времени, так и ресурсов. Для решения этой проблемы мы представляем фреймворк под названием Light Eyes или "ЛЕйес", который, в отличие от обычных фотореалистичных методов, моделирует только ключевые особенности изображения, необходимые для отслеживания глаз на основе видео, используя простое распределение света. LEyes упрощает настройку для обучения нейронных сетей в различных задачах оценки взгляда. Мы демонстрируем, что модели, обученные с использованием LEyes, неизменно соответствуют или превосходят другие современные алгоритмы с точки зрения локализации зрачков и CR в хорошо известных наборах данных. Кроме того, обученная модель LEyes превосходит стандартный в отрасли eye tracker, используя значительно более экономичное оборудование. В дальнейшем мы уверены, что LEyes произведет революцию в генерации синтетических данных для моделей оценки взгляда и приведет к значительным улучшениям видео-трекеров для отслеживания глаз следующего поколения.

# End-to-end Video Gaze Estimation via Capturing Head-face-eye Spatial-temporal Interaction Context

В этом письме мы предлагаем новый метод, взгляд с несколькими подсказками (MCGaze), для облегчения оценки взгляда на видео с помощью захвата контекста пространственно-временного взаимодействия между головой, лицом и глазами методом сквозного обучения, который еще недостаточно изучен. Основное преимущество MCGaze заключается в том, что задачи по локализации головы, лица и глаза могут быть решены совместно для оценки взгляда за один шаг с совместной оптимизацией для достижения оптимальной производительности. Во время этого происходит обмен пространственно-временным контекстом между подсказками на голове, лице и глазу. Соответственно, конечные взгляды, полученные путем объединения функций из различных запросов, могут одновременно получать глобальные подсказки от голов и лиц и локальные подсказки от глаз, что существенно повышает производительность. Между тем, одноступенчатый способ бега также обеспечивает высокую эффективность бега. Эксперименты со сложным набором данных Gaze360 подтверждают превосходство нашего предложения.

# Appearance-based gaze estimation enhanced with synthetic images using deep neural networks

Оценка взгляда человека по глазам является важным когнитивным компонентом успешного взаимодействия человека и робота, позволяя роботу считывать и предсказывать поведение человека. Мы подходим к этой проблеме с использованием искусственных нейронных сетей и создаем модульную систему оценки взгляда по отдельно обрезанным глазам, используя преимущества существующих хорошо функционирующих компонентов для распознавания лиц (RetinaFace) и оценки положения головы (6DRepNet). Предлагаемый нами метод не требует каких-либо специальных аппаратных средств или инфракрасных фильтров, но использует стандартную RGB-камеру, встроенную в ноутбук, как это часто делается с помощью методов, основанных на внешнем виде. Используя инструмент MetaHuman, мы также сгенерировали большой синтетический набор данных из более чем 57 000 человеческих лиц и сделали его общедоступным. Включение этого набора данных (с информацией о взгляде и позе головы) поверх стандартного набора данных Columbia Gaze для обучения модели привело к повышению точности при средней ошибке менее двух градусов в направлениях наклона глаз и рыскания, что выгодно отличается от аналогичных методов. Мы также проверили применимость нашей модели путем ее предварительного тестирования в реальных условиях с использованием встроенной камеры 4K в глазу полугуманоидного робота NICO.

# UVAGaze: Unsupervised 1-to-2 Views Adaptation for Gaze Estimation

Оценка взгляда стала предметом растущего интереса в недавних исследованиях. Большинство современных методов полагаются на изображения лица в одном ракурсе в качестве входных данных. Тем не менее, этим подходам сложно обрабатывать большие углы наклона головы, что приводит к потенциальным неточностям в оценке. Чтобы решить эту проблему, добавление камеры второго обзора может помочь лучше запечатлеть внешний вид глаз. Однако существующие методы с несколькими обзорами имеют два ограничения. 1) Для обучения им требуются аннотации с несколькими видами, которые являются дорогостоящими. 2) Что еще более важно, во время тестирования должны быть известны точные положения нескольких камер, которые совпадают с теми, которые использовались при обучении, что ограничивает сценарий применения. Для решения этих проблем в этой статье мы предлагаем новое решение для адаптации с 1 по 2 ракурса (1 к 2 просмотрам) - адаптивную систему с 1 по 2 ракурса без присмотра для оценки пристального взгляда (UVAGaze). Наш метод адаптирует традиционную систему оценки пристального взгляда с одним обзором для гибко размещаемых двойных камер. Здесь "гибко" означает, что мы размещаем двойные камеры в произвольных местах независимо от обучающих данных, не зная их внешних параметров. В частности, UVAGaze разрабатывает стратегию адаптации с двойным обзором и взаимным наблюдением, которая использует преимущества внутренней согласованности направлений взгляда между обоими ракурсами. Таким образом, наш метод может не только выиграть от обычной предварительной подготовки с одним взглядом, но и достичь более продвинутой оценки взгляда с двумя взглядами. Результаты эксперимента показывают, что оценщик с одним видом, адаптированный для двух видов, может достичь гораздо более высокой точности, особенно в настройках с несколькими наборами данных, со значительным улучшением на 47,0%.

# Test-Time Personalization with Meta Prompt for Gaze Estimation

Несмотря на недавние замечательные достижения в оценке взгляда, эффективная и точная персонализация оценки взгляда без меток является практической проблемой, но редко затрагивается в литературе. Для достижения эффективной персонализации мы черпаем вдохновение из последних достижений в обработке естественного языка (NLP), обновляя незначительное количество параметров, "подсказок", во время тестирования. В частности, подсказка дополнительно подключается без нарушения исходной сети и может содержать менее 1% параметров ResNet-18. Наши эксперименты показывают высокую эффективность подхода быстрой настройки. Предлагаемый метод может быть в 10 раз быстрее по скорости адаптации, чем сравниваемые методы. Однако обновить подсказку для персонализированной оценки взгляда без меток нетривиально. Во время тестирования важно убедиться, что минимизация конкретных неконтролируемых потерь приводит к достижению цели минимизации ошибки оценки взгляда. Для решения этой проблемы мы предлагаем мета-изучить подсказку, чтобы убедиться, что ее обновления соответствуют цели. Наши эксперименты показывают, что мета-выученную подсказку можно эффективно адаптировать даже при простой потере симметрии. Кроме того, мы экспериментируем с четырьмя проверками между наборами данных, чтобы показать замечательные преимущества предлагаемого метода.