

## КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ И АВТОМАТИЗАЦИЯ ПРОЕКТИРОВАНИЯ/COMPUTER MODELING AND DESIGN AUTOMATION

DOI: <https://doi.org/10.60797/IRJ.2025.161.79>

## НЕЙРОСЕТЕВАЯ МОДЕЛЬ PUPILLAI С ДВОЙНЫМ ВЫХОДОМ ДЛЯ АНАЛИЗА ЗРАЧКА В RGB-ВИДЕО

Научная статья

Кожанова М.Ю.<sup>1,\*</sup><sup>1</sup>ORCID : 0000-0002-2595-6653;<sup>1</sup>Государственный научный центр — Научно-исследовательский институт атомных реакторов, Димитровград, Российская Федерация

\* Корреспондирующий автор (ramotaugu[at]gmail.com)

**Аннотация**

В статье рассматривается задача автоматизированного выделения зрачка на изображениях глаз, полученных с помощью видеонаблюдения в видимом диапазоне спектра. В качестве обучающих данных использованы видеозаписи, полученные с применением экспериментального модуля «Айтрекинг» с установленной поворотной системой. Модуль позволяет зафиксировать положение головы и глаз относительно камеры, что упрощает обработку файлов и их последующий анализ. Для автоматизации процесса создана модель PupillAI. Модель PupillAI, основана на модифицированной архитектуре U-Net с двойным выходом: для сегментации зрачка и бинарной классификации наличия зрачка на изображении. Дополнительно реализована постобработка маски с аппроксимацией формы зрачка посредством геометрической нормализации. Для обучения модели использовали видеофайлы бинарных масок зрачков. Подготовка обучающих файлов для модели осуществлялась вручную с помощью среды Fiji. Представлены количественные и визуальные результаты, подтверждающие высокую точность модели на реальных изображениях. Геометрическая нормализация состояла в приведении геометрических данных зрачка к предпочтительной форме (круг) с максимальным сохранением положения и размеров зрачка, предсказываемых моделью. Разработка предназначена для использования в задачах психофизиологического анализа, в том числе при невозможности применения инфракрасных трекеров.

**Ключевые слова:** зрачковая сегментация, Fiji, U-Net, нейронные сети, психофизиология, бинарная классификация, постобработка.

## PUPILLAI NEURAL NETWORK MODEL WITH DUAL OUTPUT FOR PUPIL ANALYSIS IN RGB VIDEO

Research article

Kozhanova M.Y.<sup>1,\*</sup><sup>1</sup>ORCID : 0000-0002-2595-6653;<sup>1</sup>State Scientific Center — Scientific Research Institute of Nuclear Reactors, Dimitrovgrad, Russian Federation

\* Corresponding author (ramotaugu[at]gmail.com)

**Abstract**

The article examines the task of automated pupil detection in images of eyes obtained using video surveillance in the visible spectrum range. Video recordings obtained using the experimental Eye Tracking module with an installed rotary system were used as training data. The module allows the position of the head and eyes relative to the camera to be fixed, which simplifies file processing and subsequent analysis. The PupillAI model was created to automate the process. The PupillAI model is based on a modified U-Net architecture with dual output: for pupil segmentation and binary classification of the presence of the pupil in the image. Additionally, mask post-processing with pupil shape approximation by geometric normalisation was implemented. Video files of binary pupil masks were used to train the model. The training files for the model were prepared manually using the Fiji environment. Quantitative and visual results are presented, confirming the high accuracy of the model on real images. Geometric normalisation consisted of bringing the geometric data of the pupil to the preferred shape (circle) with maximum preservation of the position and size of the pupil predicted by the model. The development is intended for use in psychophysiological analysis tasks, including when infrared trackers cannot be used.

**Keywords:** pupil segmentation, Fiji, U-Net, neural networks, psychophysiology, binary classification, post-processing.

**Введение**

Зрачковая реакция является важным психофизиологическим индикатором, отражающим процессы внимания, эмоционального состояния, когнитивной нагрузки и уровня бодрствования [1], [2], [3]. Изменения диаметра зрачка и его наличие в поле зрения активно используются в нейropsychологических экспериментах, биометрии, системах адаптивного управления и интерфейсах человек-машина [4], [5], [6].

Наиболее распространённые системы айтрекинга, включая EyeLink, Pupil Labs и Tobii, предполагают использование инфракрасных камер, что требует контролируемых условий освещения, а также физического закрепления сенсоров вблизи глаза. Это ограничивает возможности их применения в полевых и мобильных условиях, а также затрудняет использование таких систем при нестабильном положении головы и при естественном освещении [7].

Альтернативой выступают решения, основанные на анализе видеопотока в видимом спектре (RGB), но существующие подходы либо требуют устойчивого фронтального положения головы, либо основаны на

синтетических или лабораторных данных [8]. Отдельные работы предлагают сегментацию зрачка на основе сверточных сетей, однако большинство моделей не учитывает случаи отсутствия зрачка в кадре (например, при моргании или частичном закрытии глаз) [9], [10].

Таким образом, актуальной задачей является разработка модели, способной:

- выполнять сегментацию зрачка на видеоданных в видимом диапазоне;
- классифицировать наличие зрачка в кадре;
- быть устойчивой к бликам, естественному освещению, частичным перекрытиям и различному положению головы.

В настоящей работе предложен подход, основанный на архитектуре U-Net с двумя выходами — сегментирующим и бинарным. Обучение модели проведено на масках подготовленных и размеченных в среде Fiji. Дополнительно реализована геометрическая нормализация масок в области зрачка, что позволило повысить точность при последующей аналитике (например, расчётах площади или центра масс).

### Методы и принципы исследования

Видеозаписи зрачковой реакции были получены с использованием модифицированной системы фиксации головы, выполненной на основе шлема виртуальной реальности с возможностью поворота по трём осям. Камера видеонаблюдения фиксировалась в постоянном положении относительно глаза, а перемещение головы осуществлялось по команде экспериментатора с контролем углов наклона. Запись осуществлялась в условиях естественного освещения, без применения инфракрасных источников.

#### 2.1. Подготовка обучающих масок в среде Fiji

В прикладных задачах машинного зрения требуется быстрый - автоматический анализ видеофайлов. Поэтому для перехода от анализа зрачковой реакции «вручную» или от полуавтоматического режима требуется использование современного инструментария. U-Net — это архитектура сверточной нейронной сети, разработанная для задач сегментации изображений, особенно в области медицинских изображений. Ее особенность — U-образная форма, образованная кодирующим и декодирующим путями, что позволяет эффективно обрабатывать изображения, сохраняя пространственную информацию. Поэтому для достижения нашей цели мы выбрали именно ее.

Как известно, для обучения модели необходим набор файлов, на которых происходит обучение. Несомненно, конечный результат обучения во многом определяется качеством обучающих файлов, т.е. в нашем случае — качеством оконтуривания. В программе Fiji представлен весь необходимый инструментарий для анализа изображений, а так же существует возможность автоматизации действий путем написания макросов. Таким образом, для подготовки набора обучающих кадров использовали программу Fiji.

Для получения видеофайлов использовали модернизированный пупиллографический модуль. Модернизация модуля обусловлена необходимостью исследования влияния угла наклона головы на точность отслеживания трека центра внимания. Жесткая координатная связь не только облегчает обработку изображений, полученных видеокамерой. Чем строже ограничения движений головы, тем меньше искажений треков, тем более однозначная интерпретация результатов. Модуль (рис. 1) дополнен блоком ручного управления пространственным положением металлического прямоугольного каркаса. Каркас приварен к стальной трубе, которую зажимают в поворотном механизме, присоединенному к вертикальной стойке. К противоположному концу стальной трубы жестко смонтирован стрелочный прибор контроля угла поворота. Выбор угла наклона осуществляется путем ослабления крепления и собственно поворотом — перемещении свободного конца стрелки вдоль шкалы транспортира до необходимого деления. Вертикальная стальная стойка позволяет устанавливать поворотный механизм на нужной высоте. Таким образом, осуществляется 3D — контроль угла поворота каркаса, к которому с внутренней стороны жестко крепится наголовный шлем от сварочной маски. Сам шлем снабжен мягкой прослойкой для комфорта испытуемого. Видеокамера крепится отдельно от шлема. Весь модуль крепится к неподвижному столу. В затылочной части наголовного шлема находится регулировочный винт, позволяющий ослабить или затянуть шлем на голове. Далее с помощью оптоэлектронной системы производится видеосъемка движения зрачка. Внутреннее время видеокамер синхронизировано до сотых секунды. Это позволяет установить элемент изображения, просмотр которого вызвал эмоциональный отклик, т.е. причину изменения размеров зрачков. На голову испытуемого надевается шлем. Поворотный механизм устанавливается на 0 (ноль) градусов. Человек, надев шлем- 3, располагается на расстоянии, при котором изменение освещенности поверхности глаза, обусловленная свечением монитора, становится незначительна. Далее с помощью оптоэлектронной системы производится видеосъемка зрачка. В экспериментах использовали видеокамеры с временным разрешением не более 33 мс, что позволяет распознавать эксплицитные (от 250 до 450 мс) и имплицитные (150–250 мс) фиксации взгляда. Для получения искаженного трека центра внимания с помощью поворотного механизма голова испытуемого поворачивается на фиксированный угол при прочих равных условиях. Общее время всей процедуры исследования занимает не более 5 минут, что не влияет на утомление респондента, но достаточно для отслеживания трека взгляда. Все видеофайлы подвергались процедурам улучшения изображения и сегментации разными методами и маскированием. По результатам статистического анализа выбирается метод, в котором отличие измерений от результатов «ручных» минимально.



Рисунок 1 - Модернизированный экспериментальный модуль «Айтрекинг»  
DOI: <https://doi.org/10.60797/IRJ.2025.161.79.1>

Для полученных изображений успешно можно применять следующий алгоритм. В первую очередь нужно обрезать файлы, так как они имеют большое разрешение, и компьютер не сможет обработать их целиком. Необходимо выделить участок изображения, содержащий зрачок. Выделение должно иметь достаточные размеры, чтобы ни на одном из кадров зрачок не покидал выделенной области. Далее выполнением команду *Subtract Background*, которая удаляет гладкие сплошные фоны. Для изображений со значениями пикселей, имеющими очень разные диапазоны, радиус “катящегося шара” должен быть обратно пропорционален диапазону значений пикселей. Данную функцию можно использовать для коррекции неравномерного освещения фона. Улучшаем контрастность для того, чтобы избавиться от дефектов: блики от источника света или тень от ресниц. После этого устанавливаем порог *Threshold*. Используя данный инструмент, можно настроить верхние и нижние значения порога для сегментирования области интереса и заднего фона изображения. Инструмент *Remove Outliers* заменяет пиксель на медиану пикселей в окружении, если он отклоняется от медианы более чем на определенное значение (порог). Процесс обработки представлен на рис. 2. Если полученный результат удовлетворительный, то изображения бинаризируют и проводят измерения.



Рисунок 2 - Этапы оконтуривания зрачка в среде Fiji  
DOI: <https://doi.org/10.60797/IRJ.2025.161.79.2>

Часто приходится проводить оконтуривание изображений, имеющих некоторые дефекты. Например, при сильном затемнении изображения вокруг зрачка после бинаризации изображения не удастся выделить зрачок, в связи с чем невозможно правильно определить его положение. При сильной зашумленности на изображении после выделения граничных точек часто не удастся верно определить радиус зрачка. Чтобы решить данную проблему, можно применить инструмент *Noise*, который может удалить шум на изображении. Если на изображении зрачка находится блик или отражение от источника света, то оконтуривание зрачка становится почти невозможным. Для избавления от тени ресниц, можно менять интенсивность пикселя: насыщенность, яркость, светлоту. Также решить данную проблему поможет применение различных цветовых настроек. Например, *Split Channels* разбивает изображение RGB на три 8-битных изображения в оттенках серого, содержащих красный, зеленый и синие компоненты оригинала. Однако, предложенные процессы занимают большое количество времени, именно поэтому автоматизация процесса оконтуривания зрачка является актуальным аспектом.

Таким образом подготовка обучающих файлов в среде Fiji включал предварительную фильтрацию (*Subtract Background*), пороговую сегментацию (*Threshold*), устранение выбросов (*Remove Outliers*) и сохранение бинарных масок с идентичной структурой имён файлов. Для каждого исходного кадра формировалась пара изображений: одно с бинарной маской зрачка (белый фон, чёрный зрачок), второе — с бликом. Все маски были приведены к унифицированному размеру (256×256 пикселей) и сгруппированы в обучающую и валидационную выборки (300 и 75 кадров соответственно).

## 2.2. Архитектура и обучение модели PupillAI

Для синхронной сегментации зрачка и определения его наличия на кадре была разработана модель PupillAI с двойным выходом, основанная на архитектуре U-Net:

- первый выход осуществляет сегментацию зрачка (выходной слой *seg\_output*, сигмоидная активация);
- второй выход представляет собой бинарную классификацию наличия зрачка (*cls\_output*, сигмоидная активация).

Модель обучалась с использованием функции потерь *binary\_crossentropy* по каждому выходу. При этом суммарная функция потерь учитывала вклад классификатора с весом 0.5. Обучение производилось в среде Google Colab с использованием оптимизатора Adam ( $\text{lr}=0.001$ ), на 30 эпохах. Использовалась библиотека TensorFlow 2.14.

После сегментации бинарная маска проходила морфологическую обработку (операция закрытия), а далее маски предсказанных форм нормировали по геометрической форме зрачка. Нормировка формы основана на неизменной форме зрачка-круг и адаптирована под текущий размер предсказанной моделью области зрачка. Это позволило наиболее точно выделять зрачок, устранить артефакты и подготовить результат для последующего количественного анализа (например, вычисления площади, центра и радиуса).

### Основные результаты

Модель PupillAI обучалась на 300 размеченных кадров с валидацией на 75 кадрах. Наилучшие показатели были достигнуты на 27-й эпохе, после чего наблюдалось насыщение по точности. Итоговые значения метрик модели PupillAI на валидационной выборке:  $\text{val\_seg\_output\_accuracy} = 0.9900$ ,  $\text{val\_cls\_output\_accuracy} = 1.0000$ ,  $\text{val\_loss} = 0.0212$ ,  $\text{val\_seg\_output\_loss} = 0.0202$ ,  $\text{val\_cls\_output\_loss} = 0.0017$ . Полученные результаты демонстрируют высокую точность как в задаче сегментации зрачка, так и в бинарной классификации наличия зрачка на кадре. Для качественной оценки модели были проведены визуальные тесты на валидационной выборке. На рисунке 3 представлены примеры кадров с соответствующими предсказаниями модели и финальной маской, слева направо: исходное изображение, истинная маска, предсказание, нормированная маска.

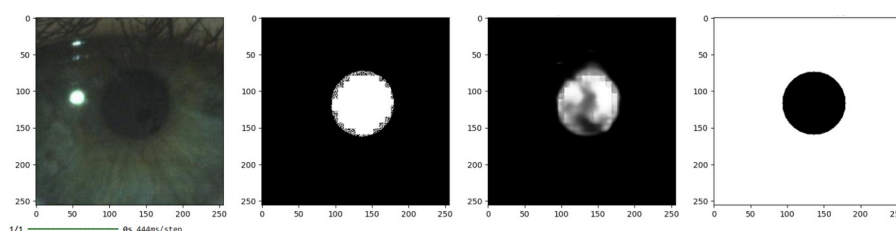


Рисунок 3 - Пример предсказания модели PupillAI  
DOI: <https://doi.org/10.60797/IRJ.2025.161.79.3>

Модель успешно сегментирует зрачок даже в условиях бликов, низкой контрастности и при частичном перекрытии глазной щели. Наличие дополнительного выходного блока `cls_output` позволяет корректно обрабатывать случаи моргания и отсутствия зрачка, исключая ошибочные маски.

В отличие от моделей, обученных на инфракрасных изображениях или синтетических данных, модель PupillAI демонстрирует устойчивую работу на RGB-изображениях без использования специализированного оборудования. Кроме того, реализация постобработки с геометрической нормализацией маски по фактической форме зрачка позволяет стандартизировать форму сегмента и минимизировать влияние шумов.

### Обсуждение

Полученные результаты подтверждают применимость предложенной модели PupillAI для задач сегментации зрачка в RGB-видеопотоке. Средняя точность по сегментационному выходу на валидационной выборке составила 0,9900, а точность бинарного выхода, отвечающего за определение наличия зрачка в кадре — 1,0000. Важно отметить, что модель корректно обрабатывает как изображения с чётким зрачком, так и кадры, где зрачок отсутствует (например, при моргании), за счёт введения второго классификационного выхода. Для 9 из 10 кадров в сложных условиях (блики, неполное открытие глаза) предсказания модели соответствовали истинным маскам с допустимым отклонением по площади не более 8%. Это позволяет применять модель в условиях, где невозможно использовать ИК-трекеры. В отличие от классических алгоритмов, таких как Swirski et al. (2012) и ExCuSe (2016), модель PupillAI не требует предварительной калибровки, устойчиво работает с нецентрированными и наклонёнными изображениями, и способна обрабатывать кадры с частичным перекрытием зрачка и бликами. Кроме того, в отличие от решений, основанных на инфракрасных данных, предлагаемая модель обучена на RGB-данных, что расширяет её применимость. Ниже приведён фрагмент архитектуры модели, демонстрирующий двойную структуру выходов:

# Выход 1: сегментация зрачка

```
seg_output = Conv2D(1, 1,
activation='sigmoid',
name="seg_output")(decoder_output)
```

# Выход 2: классификация наличия зрачка

```
cls = GlobalAveragePooling2D()
(bottleneck)
cls_output = Dense(1,
activation='sigmoid', name="cls_output")
(cls)model = Model(inputs, [seg_output,
cls_output])
```

Планируется дальнейшее дообучение модели на более сложных условиях (низкое освещение, резкие смены фокуса, перекрытие ресницами), а также переход к обработке видеопотока в реальном времени.

### Заключение

В работе предложен подход к сегментации зрачка на видеокдрах, полученных в видимом диапазоне спектра без применения инфракрасной подсветки. Используя ручную размеченные маски, подготовленные в среде Fiji, была обучена модель PupillAI на основе архитектуры U-Net с двойным выходом — для сегментации и бинарной

классификации наличия зрачка. Модель демонстрирует высокую точность на реальных данных, включая сложные случаи с бликами, частичным закрытием глаза и низкой контрастностью. Дополнительно реализован механизм постобработки результатов сегментации. Геометрическая нормализация состояла в приведении геометрических данных зрачка к предпочтительной форме (круг) с максимальным сохранением положения и размеров зрачка, предсказываемых моделью. Подобный подход позволил стандартизировать выделение зрачков, что значительно повысило эффективность и точность разрабатываемой модели.

Это особенно важно при последующей количественной оценке (площадь, центр, радиус зрачка) и потенциально может быть использовано в задачах психофизиологического анализа. Этапы архитектурного проектирования модели, подбор гиперпараметров и постобработки результатов сегментации выполнены в интерактивном взаимодействии с AI-помощником ChatGPT (OpenAI, 2025), что обеспечило гибкость и эффективность в исследовательском процессе.

Полученные результаты могут быть применимы в задачах зрачковой аналитики, встраиваемых системах видеонаблюдения, а также в исследованиях, где использование инфракрасных трекеров затруднено или невозможно.

### Конфликт интересов

Не указан.

### Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

### Conflict of Interest

None declared.

### Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

### Список литературы на английском языке / References in English

1. Akhmetvaleev A.M. Pupillogram Acquisition and Analysis Technology for Human Functional State Assessing. / A.M. Akhmetvaleev // Society 5.0: Human-Centered Society Challenges and Solutions. Cham: Springer International Publishing. — 2022. — № 416. — P. 75–86.
2. Boronenko M.P. Development of a method for determining the emotional state of a person by pupillograms. / M.P. Boronenko. // WAIE 2020: 2020 2nd International Workshop on Artificial Intelligence and Education; — New York: Association for Computing Machinery, 2020.
3. Mathôt S. Pupillometry: Psychology, physiology, and function. / S. Mathôt // Journal of Cognition 1.1. — 2018. — № 16.
4. Zemblys R. GazeNet: End-to-end eye-movement event detection with deep neural networks. / R. Zemblys // Behavior Research Methods. — 2020. — № 52 (2).
5. Koorathota S. A Recurrent Neural Network for Attenuating Non-cognitive Components of Pupil Dynamics. / S. Koorathota // Frontiers in Psychology. — 2021. — № 12.
6. Fuhl W. Excuse: Robust pupil detection in real-world scenarios / W. Fuhl, T. Kubler, K. Sippel et al. // International Conference on Computer Analysis of Images and Patterns / Ed. by G. Azzopardi, N. Petkov. — Cham: Springer, 2015. — Iss. 9256. — DOI: 10.1007/978-3-319-23192-1\_4
7. Ismagilov I. Formation of a knowledge base to analyze the issue of transport and the environment. / I. Ismagilov // Caspian Journal of Environmental Sciences. — 2020. — № 18.5. — P. 615–621.
8. Petrosyants D. Technology for collecting initial data for constructing models for assessing the functional state of a human by pupil's response to illumination changes in the solution of some problems of transport safety. / D. Petrosyants // Computer Research and Modeling. — 2021. — № 13.2. — P. 417–427.
9. Boronenko M. Pupillary Reaction as a Tool to Control the Psychophysical State of Workers / Marina Boronenko [et al.] // Human Factors in Engineering. — CRC Press, 2023. — P. 47–65.
10. Isaeva O. Application of ImageJ program for the analysis of pupil reaction in security systems. / O. Isaeva // Journal of Physics: Conference Series. — 2020. — Vol. 1519. — № 1.