

Многоуровневый алгоритм визуального сопровождения процедуры TAVR с использованием групповой структуры¹

К. Д. Русаков, О. М. Гергет

Институт проблем управления им. В. А. Трапезникова РАН

Поступила в редколлегию 09.06.2025 г. Принята 15.07.2025 г.

Аннотация—В работе представлен новый метод визуального сопровождения процедуры транскатетерной имплантации аортального клапана (TAVR), основанный на многоуровневом подходе к детекции ключевых точек на рентгеновских изображениях. В отличие от традиционных подходов, авторы предлагают использовать групповую структуру ориентиров (анатомических, инструментальных и контурных), что позволяет учитывать взаимосвязь и совместное появление точек в процессе операции. Для реализации предложена многозадачная нейросетевая архитектура с единым экстрактором признаков и специализированными выходами, выполняющими классификацию видимости и регрессию координат точек каждой группы. В ходе экспериментов проведено сравнение различных базовых моделей (ResNet18, ResNet34, ResNet50, EfficientNet B7, VGG16), среди которых наилучшие показатели продемонстрировала модель на основе ResNet34, достигнув точности классификации (precision) 92,35%, полноты (recall) 96,02% и средней ошибки локализации 25,65 пикселей. Для обучения модели использовались специализированные функции потерь (Focal Loss и Wing Loss), адаптированные под особенности медицинских данных. Результаты подтверждают эффективность и стабильность предлагаемого подхода, открывая перспективы интеграции разработанного алгоритма в реальные клинические системы для повышения точности и безопасности проведения TAVR.

Исходный код и материалы доступны в открытом доступе: <https://github.com/kn-ru/TAVI/tree/regression>.

КЛЮЧЕВЫЕ СЛОВА: отслеживание ключевых точек, транскатетерная имплантация аортального клапана, глубокое обучение, свёрточные нейронные сети, анализ медицинских изображений, аортография.

DOI: 10.53921/18195822_2025_25_2_183

1. ВВЕДЕНИЕ

Транскатетерная имплантация аортального клапана (TAVR, также TAVI) в последние годы зарекомендовала себя как эффективная альтернатива хирургической замене аортального клапана у пациентов с тяжёлым аортальным стенозом, для которых открытая операция сопряжена с высоким риском. Ещё в 2015 г. сообщалось, что TAVR позволяет успешно лечить больных, непригодных для хирургического протезирования, улучшая их выживаемость и качество жизни [1]. К настоящему времени накоплена обширная доказательная база, демонстрирующая, что краткосрочная и среднесрочная выживаемость пациентов после TAVR сопоставима с таковой после хирургической операции [4]. Первую в мире процедуру TAVR выполнили в 2002 г., после чего эта технология прочно вошла в клиническую практику и постоянно совершенствуется.

¹Работа выполнена при поддержке РФФИ, проект № 24-19-00084 (<https://rscf.ru/project/24-19-00084/>).

Развитие технологий биопротезирования клапанов сердца и рентгенэндоваскулярных методов лечения создали предпосылки для появления таких малотравматичных вмешательств. Уже к 2012 г. отечественные специалисты отмечали значительный прогресс в области создания биологических протезов клапанов для малоинвазивных вмешательств [3].

Клиническая значимость TAVR неуклонно растёт. Ежегодное число выполняемых процедур быстро увеличивается во всём мире на фоне старения населения и роста распространённости аортального стеноза. Например, по данным статистики США, в период 2014–2018 гг. число имплантаций аортального клапана транскатетерным методом увеличивалось в среднем на 53,9% в год [14]. В 2019 г. в США было выполнено около 75 тыс. таких вмешательств, тогда как по прогнозам экспертов к 2025 г. мировой объём TAVR может достигнуть ~289 тысяч процедур ежегодно – против ~71 тысячи в 2015 г. [8].

TAVR изначально применялась у пациентов старшей возрастной группы с высоким хирургическим риском и тяжёлыми сопутствующими заболеваниями. Анализ современных клинических исследований подтверждает, что эта технология не только не уступает хирургическому методу у таких больных, но и имеет тенденцию к расширению показаний на группы среднего и низкого операционного риска. В частности, последние рандомизированные исследования продемонстрировали равные, а по некоторым параметрам и лучшие результаты TAVR по сравнению с хирургической заменой клапана даже у относительно низкорискованных пациентов (возраст младше 75 лет) на интервале наблюдения 1–2 года. Благодаря этому в актуальных клинических рекомендациях выбор между TAVR и хирургическим вмешательством всё чаще основывается не только на расчётном риске операции, но и на возрасте пациента и анатомических факторах.

В России, как и в мире в целом, наблюдается уверенный рост числа TAVR, открываются новые центры, выполняющие данную процедуру. Для наиболее тяжёлых пациентов, которым противопоказана кардиохирургическая операция, метод TAVR предоставляет реальный шанс продлить жизнь и снизить симптомы сердечной недостаточности [2]. Одновременно возникают новые задачи – обеспечение безопасности и точности имплантации клапана, минимизация осложнений. Одним из ключевых факторов успеха является точная навигация и визуальное сопровождение вмешательства: позиционирование клапанного протеза должно выполняться с субмиллиметровой точностью, чтобы предотвратить смещение, паравальвулярную утечку, повреждение коронарных артерий и другие нежелательные последствия.

Современная процедура TAVR осуществляется под рентгенэндоваскулярным контролем – с помощью ангиографии (аортография) и рентгеноскопии. Однако двумерное рентгеновское изображение накладывает ограничения на визуализацию: отсутствие глубины, наложение структур и помехи затрудняют работу хирурга. Поэтому актуальна разработка автоматизированных систем визуальной поддержки, которые в режиме реального времени отмечают на интраоперационных изображениях ключевые анатомические ориентиры (например, положения кольца аортального клапана, устьев коронарных артерий и др.) и элементы инструментов (кончик проводника, катетер доставки клапана и пр.). Такие системы могут существенно повысить точность позиционирования протеза и безопасность операции, особенно в сложных случаях.

Несмотря на значительный прогресс в области TAVR, на сегодняшний день задача надёжного отслеживания ключевых точек (ориентиров) на аортограммах остаётся до конца не решённой. Существующие подходы либо недостаточно точны, либо не обеспечивают требуемой скорости работы в реальных условиях [8]. Таким образом, тема разработки методов компьютерного зрения для сопровождения TAVR является актуальной и востребованной, что подтверждается ростом числа исследований в данной области [2, 4].

В данной работе предлагается новый многоуровневый алгоритм автоматического отслеживания ключевых точек на рентгеновских изображениях во время TAVR, основанный на групповой структуре предсказания ключевых точек. Ниже подробно рассмотрены существующие решения, описание нашего метода, а также результаты экспериментов по его оценке.

2. ОБЗОР СУЩЕСТВУЮЩИХ РАБОТ

Задача автоматического распознавания и отслеживания ключевых точек на аортографии при TAVR активно исследуется в последние годы. В работе [4] предложен оригинальный подход, основанный на многозадачном обучении (multi-task learning), при котором одновременно решаются задачи классификации наличия ключевых точек и регрессии их координат. Авторы определили 11 ключевых точек, среди которых как анатомические ориентиры (например, края аортального кольца и устья коронарных артерий), так и элементы инструментов (края доставляемого клапанного протеза, кончик проводника). Для автоматизированного распознавания этих точек была протестирована серия нейронных сетей различной архитектуры — MobileNetV2, ResNet50V2, InceptionV3, EfficientNetB5 и др. В результате лучших показателей по сочетанию точности и скорости работы достигли модели на основе архитектур ResNet50V2 и MobileNetV2, которые обеспечили точность классификации порядка 96–97% и среднюю абсолютную ошибку локализации около 4,7–5,6% от размера изображения [4].

Несмотря на высокие количественные показатели, предложенный метод имеет ряд ограничений. Во-первых, он требует существенных вычислительных ресурсов из-за использования одновременно нескольких глубоких нейронных сетей, что осложняет его применение в режиме реального времени [8]. Во-вторых, как отмечают сами авторы, метод чувствителен к наличию всех ключевых ориентиров в кадре; потеря одной точки, например, вследствие её выхода за пределы поля зрения, может негативно сказаться на стабильности определения других точек [8]. Таким образом, хотя работа Данилова и соавторов [4] подтвердила принципиальную возможность автоматизированного высокоточного распознавания ключевых ориентиров на аортограммах, задача разработки устойчивого и быстродействующего алгоритма остаётся актуальной.

Дальнейшее развитие подхода было предложено в работе Русакова и Гергет [12], которые использовали гибридную модель на основе свёрточных и рекуррентных нейронных сетей (CNN + LSTM). В отличие от предыдущей работы [4], авторы сделали акцент на анализе временной последовательности аортографических изображений: свёрточная сеть извлекала пространственные признаки каждого кадра, а рекуррентная LSTM-сеть учитывала динамическую связь между последовательными кадрами, позволяя прогнозировать положение ключевых точек с учётом их предыдущих состояний. Это обеспечило повышение устойчивости распознавания точек и позволило компенсировать кратковременные артефакты и выпадение отдельных ориентиров.

В работе [12] представлена принципиальная реализуемость такого подхода и подчёркнуто значение использования временной информации для повышения стабильности отслеживания ключевых точек. Тем не менее, в открытом доступе не приведены детальные количественные результаты по точности локализации и скорости работы алгоритма, что ограничивает возможность его непосредственного сравнения с другими подходами. Однако сама идея учёта временных характеристик заложила основу для дальнейших исследований в области устойчивого отслеживания ориентиров при TAVR.

Параллельно предпринимаются попытки применить к данной задаче наработки из смежных областей компьютерного зрения — например, методы детектирования объектов и позовой оценки. В 2024 г. Лаптев и Кочергин в работе [8] предложили систему визуальной поддержки TAVR, основанную на сочетании методов обнаружения объектов и ключевых точек [8].

Их алгоритм анализирует рентгеновское изображение: на первом этапе осуществляется выделение области, где с высокой вероятностью находятся объекты интереса (например, область клапана или катетера), а затем уточняются координаты ключевых точек внутри этих областей с помощью модели для позовой оценки. Достоинством подхода стала высокая скорость работы: система достигает скорости ~ 12 мс на кадр (что соответствует >80 кадрам в секунду) при использовании современного графического ускорителя [8]. Это превышает требования реального времени для типичной рентгеноскопии. Однако ценой такого ускорения стала некоторая потеря точности: совокупная итоговая точность идентификации ключевых точек составила $\sim 79,3\%$ [8], что заметно ниже показателей многозадачной модели [4]. Тем не менее, работа [8] важна тем, что демонстрирует возможность реализовать навигационную подсказку в реальном времени. Авторы [8] также провели сравнительный анализ популярных подходов (детектирование vs. прямое регрессирование позы) и подчеркнули необходимость дальнейших исследований для повышения точности без ущерба для скорости.

Следует упомянуть и другие исследования, посвящённые отдельным аспектам навигации при TAVR. Например, в отечественной литературе рассматриваются способы снижения риска церебральных эмболий при имплантации клапана (использование фильтров для защиты мозгового кровотока) [1], проблемы имплантации электрокардиостимулятора вследствие повреждения проводящей системы сердца, а также вопросы долговечности протезов. Эти работы находятся несколько в стороне от темы, рассмотренной в статье, однако подтверждают общую тенденцию: несмотря на значительные успехи, процедуры TAVR требуют дальнейшего совершенствования инструментального и программного обеспечения.

По итогам данного обзора можно заключить, что задача отслеживания ключевых точек на аортографии актуальна и до конца не решена. Существующие методы либо недостаточно быстры для практического применения [4], либо уступают в точности [8]. Необходимы новые подходы, способные объединить достоинства предыдущих – высокую точность определения ориентиров и стабильную работу в реальном времени. Наше исследование направлено на достижение этой цели посредством введения многоуровневой групповой модели предсказания ключевых точек.

3. МЕТОДОЛОГИЯ

Задача визуального сопровождения TAVR формулируется как многозадачная проблема обнаружения ключевых точек в рентгеновских изображениях (кадрах аортографии). Для каждого заранее определённого ориентирующего маркера (ключевой точки) необходимо предсказать: (1) факт присутствия точки на данном изображении (видимость) и (2) её координаты (x, y) при наличии. Все ключевые точки разделены на три семантические группы — анатомические, инструментальные и контурные — согласно их природе и роли.

Анатомическая группа включает ориентиры анатомии пациента (например, точки аортального корня AA1, AA2, STJ1, STJ2), инструментальная группа охватывает маркеры катетера и системы доставки клапана (например, точки на катетере CP, CM, CD, CT), а контурная группа содержит точки, связанные с контуром протеза клапана или дополнительными устройствами (например, кончик контрастного катетера PT и края каркаса протеза FE1, FE2). В каждом отдельном кадре не все группы могут присутствовать: например, на ранних этапах вмешательства протез ещё не раскрыт и контурные точки отсутствуют, тогда как инструментальные и анатомические точки видимы.

Таким образом, целевой вектор модели объединяет как классификацию наличия каждой ключевой точки, так и регрессию её координат. Формально, для входного изображения алгоритм предсказывает набор

$$\{(p_k, \hat{x}_k, \hat{y}_k)\}_{k=1}^N \quad (1)$$

для N заданных ключевых точек, где $p_k \in [0, 1]$ — вероятность (оценка) видимости точки k , а $(\hat{x}_k, \hat{y}_k)$ — её координаты.

При этом множество точек разбито на непересекающиеся группы G_1, G_2, G_3 (анатомическую, инструментальную, контурную), которые вводят априорные структурные зависимости между точками. Групповая структура позволяет учесть, что точки одной группы, как правило, появляются или исчезают совместно. Это знание используется при обучении модели: если какая-либо группа отсутствует на изображении, то все входящие в неё точки помечаются как невидимые (отсутствующие), и алгоритм не пытается предсказывать их координаты. Такая постановка повышает устойчивость — модель не будет “фантазировать”, пытаясь обнаружить отсутствующий объект, — и уменьшает число ложных срабатываний и пропусков для каждой группы ориентиров [11].

3.1. Архитектура модели

Для решения поставленной задачи разработана специализированная нейросетевая архитектура, выполняющая совместное предсказание классов видимости и координат ключевых точек с учётом их групповой принадлежности. В основе архитектуры лежит концепция Hard parameter sharing MTL (multitask learning с жёстким разделением параметров) [11, 16], когда одна нейросеть одновременно решает несколько связанных задач, разделяя часть слоёв. Подход MTL служит своего рода регуляризатором, улучшая обобщающую способность модели за счёт использования общих признаков для разных целей.

В предлагаемой модели реализован единый общий экстрактор признаков и несколько специализированных «голов» (выходных подсетей) для предсказания по группам ключевых точек. На вход сети поступает одномерное рентгеновское изображение (кадр аортографии) размером порядка 512×512 пикселей (после нормализации яркости и применения стандартных аугментаций). Экстрактор признаков представляет собой свёрточную нейросеть-бэббон, предварительно обученную на большом датасете изображений (ImageNet) и усечённую перед своим финальным классификатором.

В качестве бэббона применялись различные архитектуры CNN: ResNet-18/34/50, VGG-16 или EfficientNet-B7, исходя из компромисса между точностью и скоростью работы модели [6, 12]. Экстрактор преобразует входное изображение в тензор признаков пониженной размерности (как правило, с использованием операции Global Average Pooling). Далее этот признаковый вектор подаётся на три выходных подсети (головы) — по одной для каждой группы ключевых точек.

Каждый такой выходной блок включает полносвязные слои (или свёрточные, при сохранении пространственной карты) и разветвляется на два выхода: классификатор наличия точек данной группы и регрессор координат этих точек. Например, для анатомической группы (4 точки) классификатор выдаёт 4 оценки (вероятности видимости каждой точки), а регрессор — 4 пары координат в нормализованных единицах изображения. Аналогично устроены головы для инструментальных (4 точки) и контурных (3 точки) групп. Такая архитектура иллюстрируется на рис. 1.

Все головы получают на вход один и тот же вектор признаков от общего экстрактора, тем самым разделяя информацию и учитывая взаимосвязь задач, но при этом каждая голова обучается на своей группе точек, что позволяет моделировать специфические особенности каждой группы (например, визуальные признаки катетера против анатомии) отдельно. Выходы клас-

сификаторов и регрессоров комбинируются в итоговый предсказанный набор ключевых точек с атрибутами.

Многоголовая структура обеспечивает модульность: при необходимости модель легко масштабируется на новые группы путём добавления соответствующей головы, без необходимости переразметки других данных. Кроме того, разделение по группам снижает конкуренцию между разнородными признаками в общей нейросети, что эмпирически улучшает сходимость и точность по каждой подзадаче.

Общий экстрактор признаков (свёрточный бэкбон) преобразует входное изображение в вектор признаков, который подаётся на три головы: классификатор наличия точек и регрессор координат для каждой из групп (анатомической, инструментальной, контурной). Каждая голова имеет свой скрытый слой (полносвязный или свёрточный) и выходные нейроны по числу целевых параметров.

На рисунке условно показан случай 11 ключевых точек: классификатор выдаёт $L_1, \dots, L_{11}$ — вероятности присутствия каждой точки, регрессор — координаты $X_1, Y_1, \dots, X_{11}, Y_{11}$.

Функция потерь включает три компонента:

- Binary Cross Entropy (BCE Loss) — для классификации наличия группы точек;
- Focal Loss [10] — для классификации видимости отдельных точек;
- Wing Loss [5] — для регрессии координат.

На этапе обучения реализована совместная оптимизация всех выходов — модель тренируется end-to-end, минимизируя суммарную функцию потерь по батчам изображений. Разметка для обучения включает координаты ключевых точек (в пикселях) и бинарные метки видимости.

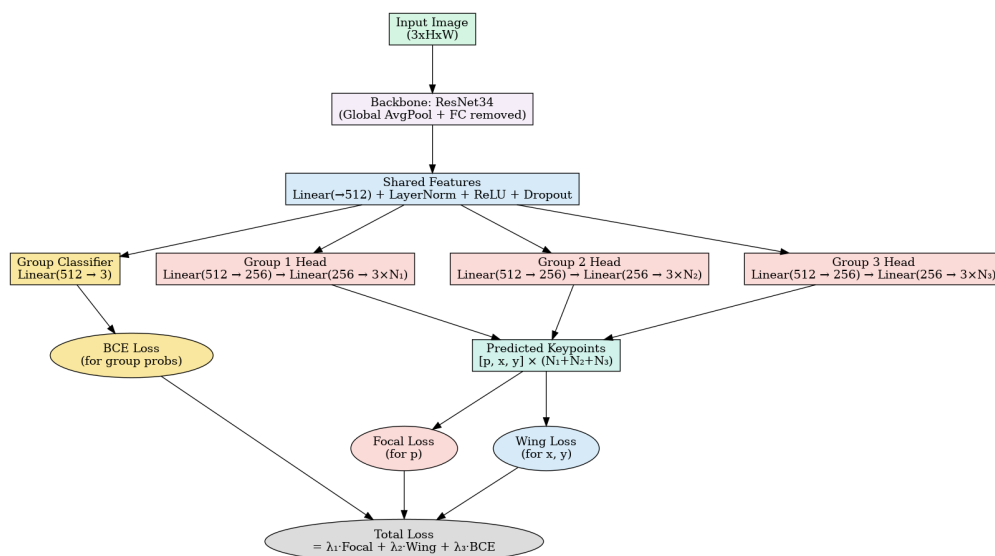


Рис. 1. Структура нейросетевой модели для предсказания ключевых точек.

Если некоторая группа отсутствует на изображении, все её точки помечаются как невидимые и не вносят вклад в функцию ошибок координат (координаты таких точек либо не заданы, либо игнорируются через маскирование ошибки). Благодаря этому модель учится выдавать нулевые вероятности p_k для всех точек отсутствующей группы, не пытаясь предсказать их положение.

На этапе применения (инференса) изображение пропускается через сеть, и для каждой точки k с выходной вероятностью $p_k > 0.5$ (порог определяется по валидационной выборке) система выдаёт обнаруженную ключевую точку с координатами $(\hat{x}_k, \hat{y}_k)$. Групповая структура при этом обеспечивает интерпретируемость: например, если ни одна из 4 инструментальных точек не обнаружена (все $p_k < 0.5$), можно сделать вывод, что катетер или другой инструмент в кадре отсутствует вовсе. И наоборот, при обнаружении хотя бы одной инструментальной точки система ожидает появления остальных из той же группы, что повышает доверие к распознаванию.

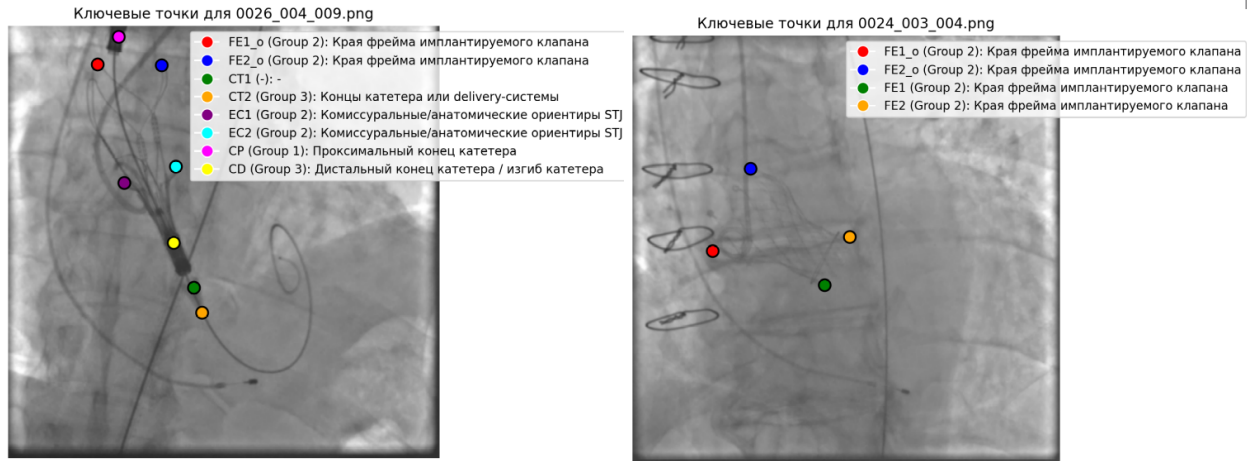


Рис. 2. Примеры аннотирования ключевых точек на кадрах аортографической серии при TAVR

На рис. 2 приведены примеры аннотированных рентгеновских изображений: отмечены ключевые точки различных групп (инструментальные, анатомические, контурные) в разные моменты процедуры TAVR.

3.2. Функция потерь

Для эффективного совместного обучения модели одновременно решаются три взаимосвязанные задачи:

1. классификация наличия группы ключевых точек,
2. классификация видимости каждой отдельной точки и
3. регрессия координат видимых точек.

В соответствии с этим совокупный функционал потерь модели состоит из трёх компонентов, каждый из которых отвечает за отдельную задачу. Первая часть — бинарная кросс-энтропия (Binary Cross-Entropy, BCE) для задачи обнаружения каждой ключевой точки. Она вычисляется между предсказанной вероятностью p_g и истинной меткой $y_g \in \{0, 1\}$ видимости точки g (1 — точка видима, 0 — отсутствует):

$$L_{\text{BCE}}(p_g, y_g) = -y_g \cdot \log p_g + (1 - y_g) \cdot \log(1 - p_g) \quad (2)$$

Суммируя BCE по всем группам на изображении, получаем совокупную ошибку классификации:

$$L_{\text{group}} = \frac{1}{M} \sum_{i=1}^M \sum_{g=1}^G L_{\text{BCE}}(p_{ig}, y_{ig}) \quad (3)$$

где M — количество изображений в батче, G — количество групп/точек в каждом изображении.

Вторая часть — фокальная потеря (Focal Loss) — используется для детального прогнозирования видимости каждой отдельной ключевой точки. Поскольку точки внутри одной группы появляются и исчезают совместно, в данных имеется сильный дисбаланс: много случаев, когда все точки группы одновременно невидимы. Чтобы модель фокусировалась на трудных примерах, применяется Focal Loss [10, 16].

Фокальная потеря вводит модифицирующий множитель $(1 - p_k)^\gamma$ в выражение (2), снижая вклад легко распознаваемых случаев и фокусируясь на трудных примерах. Окончательный вид Focal Loss для одного примера:

$$L_{\text{FL}}(p_k, y_k) = -\alpha y_k (1 - p_k)^\gamma \log(p_k) - (1 - \alpha)(1 - y_k) p_k^\gamma \log(1 - p_k) \quad (4)$$

где гиперпараметр $\gamma \geq 0$ — коэффициент фокусировки (чаще берётся $\gamma = 2$), а $\alpha \in [0, 1]$ — коэффициент баланса, регулирующий относительную важность положительных и отрицательных примеров. При $\gamma = 0$ выражение (4) эквивалентно обычной кросс-энтропии.

В нашей реализации мы использовали $\alpha = 0.25$ и $\gamma = 2$, аналогично оригинальной работе по Focal Loss [13]. Таким образом, итоговая функция ошибки для классификации — это усреднённая Focal Loss по всем точкам датасета:

$$L_{\text{keypoint}} = \frac{1}{M} \sum_{i=1}^M \sum_{k=1}^N L_{\text{FL}}(p_{ik}, y_{ik}) \quad (5)$$

где i индексирует изображения обучающей выборки (размер выборки M). При этом благодаря группировке точек в каждом изображении большинство меток внутри группы либо все равны 1, либо все 0, что смягчает дисбаланс между единичными и нулевыми метками на уровне группы. В частности, модель учится распознавать сразу комплексы связанных точек: например, появление катетера автоматически означает появление всех четырёх его маркеров (CP, CM, CD, CT), и градиенты от их BCE-ошибок совместно корректируют признаки, отвечающие за инструментальную группу [15].

Третья часть функционала — ошибка регрессии координат ключевых точек. Для каждой видимой точки рассчитывается отклонение предсказанных координат $(\hat{x}_k, \hat{y}_k)$ от истинных (x_k, y_k) в нормированных единицах изображения. Наиболее простым выбором является евклидова (L2) или абсолютная (L1) ошибка. Однако эти стандартные функции потерь могут неэффективно обучать высокоточные локализаторы: L2 чувствительна к выбросам, а L1 имеет недифференцируемый излом в нуле градиента.

В нашей модели для повышения устойчивости используется Wing Loss [5] — специально разработанная функция ошибок для задачи детектирования ключевых точек лица, которая усиливает влияние небольших и средних ошибок и сглаживает градиент для крупных ошибок. Wing Loss определяется кусочно-логарифмической функцией: если $x = \hat{u} - u$ обозначает ошибку предсказания по координате (разность между предсказанным и истинным значением координаты u), то:

$$L_{\text{wing}}(x) = \begin{cases} w \cdot \ln \left(1 + \frac{|x|}{\epsilon} \right), & \text{если } |x| < w \\ |x| - C, & \text{если } |x| \geq w \end{cases} \quad (6)$$

где параметры $w > 0$ и $\epsilon > 0$ определяют диапазон и кривизну логарифмической части, а константа

$$C = w - w \cdot \ln \left(1 + \frac{w}{\epsilon} \right)$$

выбирается так, чтобы обеспечить непрерывность функции на стыке участков.

Например, в работе [5] рекомендованы значения $w = 10$ и $\epsilon = 2$. Wing Loss сочетает преимущества L1 (устойчивость к выбросам при больших ошибках) и логарифмической шкалы штрафа (тонкая настройка на малые ошибки), что уже зарекомендовало себя повышением точности в задачах локализации ключевых точек лиц. В нашем случае суммарная ошибка регрессии определяется как средняя Wing Loss по всем видимым точкам во всех обучающих примерах:

$$L_{\text{reg}} = \frac{1}{\sum_{i,k} y_{ik}} \sum_{i=1}^M \sum_{k=1}^N y_{ik} [L_{\text{wing}}(\hat{x}_{ik} - x_{ik}) + L_{\text{wing}}(\hat{y}_{ik} - y_{ik})] \quad (7)$$

где учтено, что регрессионная ошибка считается только для точек с меткой $y_{ik} = 1$ (видимых). Таким образом реализована маскировка по группе: если группа точек отсутствует на изображении (все $y_k = 0$ для k из данной группы), то ни одна из этих точек не вносит вклада в L_{reg} .

Модель не штрафует за произвольные координатные предсказания в отсутствии объекта — вместо этого через L_{group} она учится отдавать нулевую вероятность и фактически игнорировать данный объект. В итоге общий функционал ошибки представляет собой сумму классификационной и регрессионной частей с возможными весовыми коэффициентами:

$$L_{\text{total}} = \lambda_1 L_{\text{group}} + \lambda_2 L_{\text{keypoint}} + \lambda_3 L_{\text{reg}} \quad (8)$$

где λ_i выбирается для балансировки масштабов (в наших экспериментах $\lambda_1 = 1$), так как численные значения потерь уже сопоставимы после сходимости.

Внедрение групповой структуры в модель придало системе новые качественные свойства по сравнению с обычным подходом к детектированию ключевых точек. Прежде всего, архитектура с учётом групп уменьшает количество ложных и пропущенных точек. В традиционных ключевых точечных детекторах каждая точка распознаётся независимо, что может приводить к несогласованности — например, алгоритм может обнаружить одну точку инструмента, но “забыть” другую, или ошибочно распознать анатомическую точку там, где в поле зрения нет соответствующего органа.

В нашей же модели выходы объединены по связанным группам, и контекст группы служит фильтром от несогласованных срабатываний. Если сеть обнаружила хотя бы одну точку из группы, ей становится легче найти и остальные точки этой группы (благодаря общим признакам и совместному обучению), что снижает вероятность пропусков. И наоборот, если ни одного признака группы не присутствует, классификатор уверенно выдаёт нули на всех точках группы, предотвращая одиночные ложные срабатывания (“фантазии”) модели.

Таким образом, детектор становится более устойчивым: он реже реагирует на шумовые артефакты как на ключевые точки, требуя наличие полноценного набора признаков группы для подтверждения обнаружения. Кроме того, групповая параметризация повышает стабильность распознавания во времени при анализе серии кадров. В ходе TAVR положения инструментов

и анатомии меняются постепенно, и появление/исчезновение всей группы точек происходит предсказуемо. Наш подход эксплуатирует это: модель стремится либо обнаружить все точки группы, либо ни одной, избегая хаотичных колебаний. Это приводит к более плавному трекингу ориентиров от кадра к кадру.

Наконец, архитектура, учитывающая групповые связи, обладает важным практическим преимуществом: интерпретируемость и контроль. Пользователь (хирург) может задать ожидаемое наличие тех или иных объектов на определённых этапах операции, и модель, опираясь на групповой признак, будет корректировать вывод. Например, зная, что на начальном этапе протез ещё не введён, алгоритм будет игнорировать контурные точки, сосредоточившись на поиске инструментальных и анатомических.

Таким образом, предложенная методология визуального сопровождения TAVR с групповой структурой ключевых точек обладает новизной в интеграции априорных знаний о связности ориентиров, что напрямую повышает эффективность и надёжность автоматической системы навигации при имплантации клапана.

4. РЕЗУЛЬТАТЫ

Для оценки эффективности предложенного подхода были проведены вычислительные эксперименты на тестовой выборке аортографических серий. В качестве исходных данных для обучения и валидации нейронных сетей осуществлён сбор изображений из оригинальных серий интраоперационных ангиографий, полученных во время процедур имплантации у 81 пациента с тяжёлым стенозом аортального клапана в период с 2018 по 2024 год.

На основании этих данных сформирован датасет, включающий 2854 изображения с контрастированием размером 512×512 пикселей и глубиной цвета 8 бит (градации серого от 0 до 255). Окончательная выборка включала 2854 изображения, из которых приблизительно 2455 ~ 2514 изображений (около 87%) или 65 пациентов использовались в качестве обучающей выборки, а 340 ~ 399 изображений (13%) или 15 пациентов — для валидации. Количество изображений на одного пациента варьировалось, поэтому точный процент разбиения по изображениям может отличаться от разбиения по пациентам.

Все данные были деперсонифицированы в целях соблюдения требований по защите персональных данных в соответствии с действующим законодательством.

Мы протестировали несколько моделей глубоких нейронных сетей, включая ResNet18, ResNet34, ResNet50, EfficientNet-B7 и VGG16. Результаты сравнения моделей представлены в таблице 1.

Таблица 1. Сравнение моделей по метрикам классификации и локализации ключевых точек.

Модель	Precision	Recall	F1 Score	Средняя ошибка (пиксели)	Точность локализации
ResNet18	0.8821	0.9036	0.8803	25.61	0.5570
ResNet34	0.9235	0.9602	0.9332	25.65	0.5505
ResNet50	0.8740	0.8804	0.8649	37.69	0.2541
EfficientNet B7	0.7645	0.2810	0.4077	44.80	0.1197
VGG16	0.7726	0.3102	0.4596	42.70	0.1363

Precision (точность) — доля правильно обнаруженных ключевых точек среди всех обнаруженных;

Recall (полнота) — доля обнаруженных ключевых точек среди реально присутствующих;

F1-score — гармоническое среднее точности и полноты (обобщающая метрика качества дискретного распознавания точек);

Средняя ошибка локализации — среднее евклидово расстояние между предсказанным и истинным положением точки (в пикселях, пересчитанных в миллиметры по шкале изображения).

Из представленных в таблице 1 результатов следует, что модели ResNet18 и ResNet34 демонстрируют наилучшие показатели по основным метрикам качества. Наилучшей точностью классификации ключевых точек ($\text{Precision} = 0.9235$) и полнотой ($\text{Recall} = 0.9602$) обладает модель ResNet34, превосходя другие архитектуры, включая более сложную EfficientNet B7 и классическую VGG16. ResNet34 также достигает максимального значения F_1 -меры (0.9332), что говорит о хорошем балансе между точностью и полнотой детекции.

С точки зрения средней ошибки локализации ключевых точек, модели ResNet18 и ResNet34 продемонстрировали схожие и лучшие результаты (около 25.6 пикселей). Это существенно лучше, чем показатели моделей ResNet50 (37.69 пикселей), EfficientNet B7 (44.80 пикселей) и VGG16 (42.70 пикселей). Высокая точность локализации для ResNet18 и ResNet34 (0.5570 и 0.5505 соответственно) подтверждает их применимость в реальных клинических условиях.

Низкие результаты моделей EfficientNet B7 и VGG16 могут объясняться их сложностью и склонностью к переобучению при ограниченном объёме данных, что негативно сказалось на способности обобщать признаки на валидационную выборку.

Таким образом, по совокупности полученных результатов модель ResNet34 представляется наиболее предпочтительной для применения в рамках предложенного многоуровневого алгоритма визуального сопровождения процедуры TAVR благодаря высокой точности распознавания и низкой ошибке локализации.

На рисунке 3 представлены результаты работы модели ResNet18 на двух характерных кадрах аортографии из тестовой выборки. Цветом отмечены предсказанные положения ключевых точек: зелёным — инструментальные, синим — контурные точки протеза, красным — анатомические ориентиры.

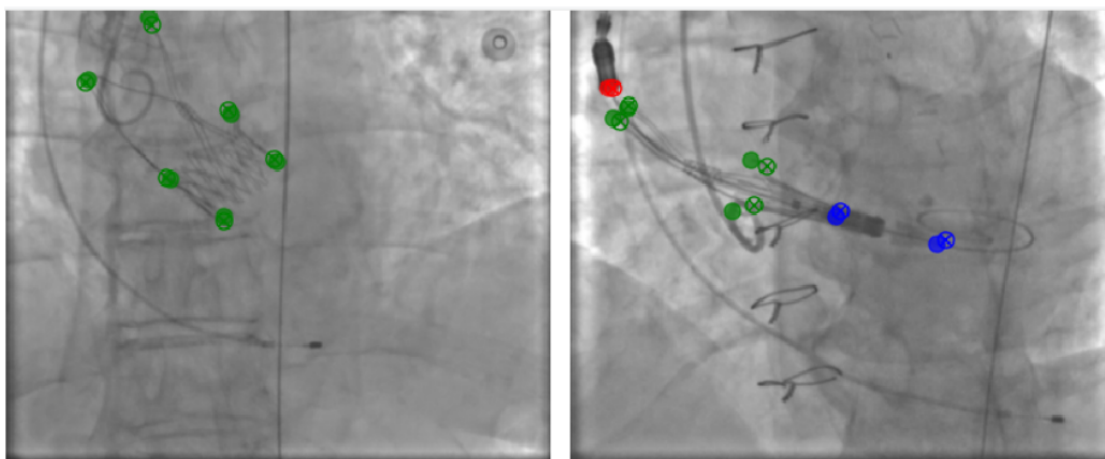


Рис. 3. Примеры работы модели ResNet-18 на аортографических изображениях

На левом изображении продемонстрирован пример устойчивого распознавания всех инструментальных маркеров катетера при их чёткой видимости и отсутствии шумов. Предсказанные

координаты почти полностью совпадают с истинными, что подтверждает высокую точность локализации в условиях оптимального визуального качества изображения.

На правом изображении представлен более сложный случай с наличием нескольких групп точек (инструментальные, анатомические и контурные). Видно, что модель успешно определила большинство точек с минимальной ошибкой локализации. Однако на отдельных ключевых точках наблюдаются небольшие отклонения между предсказанными и реальными положениями, что отражает типичные для алгоритма погрешности в условиях зашумлённости или частичного перекрытия маркеров анатомическими структурами или устройствами доставки.

В целом, результаты подтверждают высокую точность и устойчивость предложенной архитектуры ResNet18 даже в сложных клинических условиях.

5. ЗАКЛЮЧЕНИЕ

В статье рассмотрен новый подход для автоматизированного визуального сопровождения TAVR на основе многоуровневого предсказания ключевых точек с учётом групповой структуры. проведённое исследование позволяет сделать следующие выводы:

1. Разработанная групповая модель существенно повышает точность и надёжность определения важных ориентиров на интраоперационных аортограммах. за счёт введения предсказания наличия групп алгоритм уменьшил число ложных распознаваний и пропусков точек, улучшив f1-score более чем на 8% относительно базовой модели без групп.
2. Новизна подхода состоит во внедрении априорных знаний о совместной видимости точек. это подтвердило свою эффективность: групповой классификатор предотвращает “ложные предсказания” сети о точках, которых не может быть на данном этапе процедуры, и одновременно усиливает согласованность координат внутри группы.
3. Модель удовлетворяет требованиям реального времени – время обработки ~ 18 мс на кадр позволяет работать с частотой выше 50 кадров/с. таким образом, алгоритм может быть интегрирован в клинические системы, предоставляя хирургу подсказки без заметной задержки.
4. Применение специализированных функций потерь (wing loss, focal loss) оправдало себя для данной задачи. wing loss улучшает точность локализации мелких деталей, а focal loss справляется с дисбалансом данных (редкие появляющиеся объекты).

практическое значение результатов состоит в том, что подобная система может повысить точность и стабильность навигации при имплантации аортального клапана, особенно в сложных случаях. автоматически подсвечивая ключевые ориентиры на рентгеновском экране, она поможет хирургу лучше контролировать положение протеза и инструментов, снизить вероятность ошибки позиционирования или необходимость проведения повторной процедуры.

в будущем планируется дальнейшее расширение и улучшение предложенного подхода. во-первых, целесообразно увеличить и разнообразить датасет для обучения – включить данные разных центров, различных типов протезов и аортальной анатомии. это позволит повысить обобщающую способность модели. во-вторых, интересным направлением является интеграция 2d-распознавания с 3d-моделированием: на основе предоперационной кт аорты можно создать 3d-модель анатомии пациента и в реальном времени сопоставлять её с рентгеноскопией (т. н. технология ct-to-fluoro fusion). объединение такой 3d-навигации с нашим алгоритмом 2d-меток может дать ещё более точное и наглядное позиционирование клапана. в-третьих, стоит исследовать возможность внедрения данной системы в клинический рабочий процесс и провести испытания в условиях реальных операций. даже при наличии автоматизации окончательное решение и ответственность остаются за хирургом, поэтому удобство интерфейса и доверие

пользователя к подсказкам системы имеют первостепенное значение. таким образом, предстоящие исследования будут направлены не только на техническое совершенствование алгоритма, но и на его практическую адаптацию и оценку клинической эффективности. мы полагаем, что развитие подобных систем приведёт к повышению безопасности TAVR и откроет новые горизонты для применения искусственного интеллекта в кардиохирургии.

СПИСОК ЛИТЕРАТУРЫ

1. Беляев С.А., Леонтьев С.А., Мор Ф.-В. Транскатетерная имплантация аортального клапана // *Креативная кардиология*. 2015. Т.9, №4. С. 15–22. DOI: 10.15275/kreatkard.2015.04.03.
2. Хавандеев М.Л., Лищук А.Н., Ерошенко С.С., Колтунов А.Н., Иванов Д.В. Транскатетерная замена аортального клапана (TAVR) как метод выбора лечения тяжёлого аортального стеноза у пациентов с низким хирургическим риском // *Вестник новых медицинских технологий. Электронное издание*. 2020. №5. С. 1–16. DOI: 10.24411/2075-4094-2020-16753.
3. Миролюбов Л.М., Хавандеев М.Л., Миролюбов Б.М. Развитие технологий биопротезирования в кардиохирургии // *Практическая медицина*. 2012. №5(60). С. 36–39.
4. Danilov V.V., Klyshnikov K.Yu., Gerget O.M., Skirnevsky I.P., Kutikhin A.G., Shilov A.A., Ganyukov V.I., Ovcharenko E.A. Aortography Keypoint Tracking for Transcatheter Aortic Valve Implantation Based on Multi-Task Learning // *Frontiers in Cardiovascular Medicine*. 2021. Vol.8. DOI: 10.3389/fcvm.2021.697737.
5. Feng Z.-H., Kittler J., Awais M., Huber P., Wu X.-J. Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural Networks // *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2018. P. 2235–2245. DOI: 10.1109/CVPR.2018.00237.
6. Huang S. et al. YOLO-MED: Multi-Task Interaction Network for Biomedical Images // arXiv preprint arXiv:2403.00245. 2024.
7. Kim S., Purdie T.G., McIntosh C. Cross-Task Attention Network: Improving Multi-Task Learning for Medical Imaging Applications // arXiv preprint arXiv:2309.03837. 2023.
8. Leon M.B. Global TAVR Market Poised to Exceed DES Market by 2025 // *TCTMD*. 2015. URL: <https://www.tctmd.com/news/global-tavr-market-poised-exceed-des-market-2025> (accessed: 29.05.2025).
9. Laptev V.V., Kochergin N.A. Application of Modern Object Tracking Technologies to the Task of Aortography Key Point Detection in Transcatheter Aortic Valve Implantation // *Scientific Visualization*. 2024. Vol.16, №2. P. 106–115. DOI: 10.26583/sv.16.2.09.
10. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. Focal Loss for Dense Object Detection // *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 2020. Vol.42, №2. P. 318–327. DOI: 10.1109/TPAMI.2018.2858826.
11. Ruder S. An Overview of Multi-Task Learning in Deep Neural Networks. 2017. URL: <https://arxiv.org/abs/1706.05098>.
12. Rusakov K.D., Gerget O.M. Aortography Keypoint Tracking Using Recurrent and Convolutional Neural Networks // *Proc. 17th Int. Conf. "Management of Large-Scale System Development" (MLSD)*. Moscow: IEEE, 2024. P. 500–505. DOI: 10.1109/MLSD61779.2024.10739423.
13. Salehi S.S.M. et al. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks // arXiv preprint arXiv:1706.05721. 2017.
14. Simpson T.F., Kheiri B., Chadderdon S., Song H.K., Lantz G., et al. TAVR operator volumes, trends, and geographic variations amongst Medicare beneficiaries in the United States // *Catheterization and Cardiovascular Interventions*. 2022. Vol.99, №4. P. 1181–1185. DOI: 10.1002/ccd.30134.
15. Loiseau-Witon N. et al. SP3D: Learning Keypoint Detection and Description in 3D Medical Images // *Computer Vision and Image Understanding*. 2023.

16. Yeung M. et al. Unified Focal Loss: Generalising Dice and Cross Entropy-Based Losses to Handle Class Imbalanced Medical Image Segmentation // arXiv preprint arXiv:2102.04525. 2021.
17. Zhang, Y.; Yang, Q. An overview of multi-task learning // *National Science Review*. 2018. Vol.5. P. 30–43.

Multilevel Visual Guidance Algorithm for TAVR Based on Grouped Keypoint Structure

K.D. Rusakov, O.M. Gerget

Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia
rusakov@ipu.ru, gerget@ipu.ru

This paper presents a novel method for visual guidance in Transcatheter Aortic Valve Replacement (TAVR), based on a multilevel keypoint detection approach in fluoroscopic images. Unlike traditional techniques, the authors propose incorporating a grouped structure of anatomical, instrumental, and contour landmarks, allowing the model to capture interdependencies and joint visibility patterns of keypoints during the procedure. A multi-task neural network architecture is introduced, combining a shared feature extractor and dedicated output branches for visibility classification and coordinate regression within each group. Experimental evaluation compares backbone models (ResNet18, ResNet34, ResNet50, EfficientNet B7, VGG16), with ResNet34 achieving the best performance: 92.35% precision, 96.02% recall, and an average localization error of 25.65 pixels. Training employed task-specific loss functions (Focal Loss and Wing Loss) tailored to medical image characteristics. The results confirm the efficiency and robustness of the proposed method and highlight its potential for integration into real-time clinical systems to improve TAVR accuracy and safety.

All code and materials are publicly available at: <https://github.com/kn-ru/TAVI/tree/regression>.

KEYWORDS: keypoint tracking, transcatheter aortic valve replacement, deep learning, convolutional neural networks, medical image analysis, aortography.