

2. Обзор литературы

2.1 Постановка задачи и ключевые сложности.

Задача автоматической классификации архитектурных стилей по визуальным данным лежит на стыке архитектуроведения и компьютерного зрения. В наиболее общем виде она формулируется как отображение изображения здания (его фасада) в дискретное пространство стилевых категорий: исторических эпох (романтизм, готика, барокко, модерн и т.д.), региональных традиций (мексиканская колониальная архитектура, китайская традиционная, европейский модернизм) или авторских почерков отдельных архитекторов.

Практика и теория архитектуры подчеркивают, что стиль определяется совокупностью признаков: формы и композиции объёма, структурной схемы, системы деталей, материалов и цветовой палитры, а также связи здания с контекстом. В работах по глубокому обучению эта многомерность постепенно начинает учитываться: от первых подходов, где стиль описывался глобальной геометрией фасада, к более сложным моделям, анализирующим отдельные фасадные элементы, материалы, цвет и даже текстовые описания.

Ряд современных исследований явно формулируют ключевые трудности задач стилиевой классификации. В работе Li и Dong, посвященной классификации архитектурных стилей на основе глубинных моделей и CAD-данных, выделяются: разнообразие всех возможных стилей, множественность влияющих факторов (история, культура, география), субъективность экспертной атрибуции, сложность сбора и разметки репрезентативных данных, технические ограничения существующих алгоритмов и необходимость междисциплинарной интеграции архитектурной теории и вычислительных методов. Эти тезисы во многом универсальны и задают контекст для рассматриваемых далее работ.

Литература ниже структурирована по типу используемых данных и методов: от классических признаков и моделей, формируемых через сверточные сети - к семантической сегментации фасадов, анализу материалов и цвета, интеграции с CAD/BIM и, наконец, к подходам на основе vision-language моделей, ориентированным на интерпретируемый анализ стиля.

2.2. Датасеты и постановки задач.

Существенной предпосылкой развития методов классификации стилей стало появление специализированных датасетов.

Одним из первых крупных наборов данных является датасет архитектурных стилей, предложенный Xu и соавт. В нём содержится около 5000 изображений фасадов, собранных из Wikimedia Commons и отнесенных к 25 стилям - от древнеримской и готической архитектуры до современных направлений (International Style, Postmodernism, Deconstructivism и др.). Каждый стиль представлен от 50 до 300 изображений.

На базе этого датасета позже строятся более сложные методы (Zhao и др.), поэтому он важен как эталон для классических подходов. Однако он ограничен фасадами преимущественно западной архитектуры, использует плоский список стилей и опирается на web-разметку, подверженную шуму.

Отдельную линию представляют региональные наборы данных. Montoya Obeso и соавт. формируют датасет мексиканской архитектуры, разделенный на три стилистические категории: доколониальная (prehispanic), колониальная и современная, плюс класс «other» для нерелевантных изображений. Коллекция получена из видеоматериалов культурных учреждений Мексики, кадры вручную отнесены к классам. На каждый из четырех классов выделяется по 2000 изображений для обучения и 2000 для валидации, а затем набор расширяется с помощью аугментаций (повороты, зеркальные отражения) до 64000 обучающих примеров. Это один из первых датасетов, где стили формально заданы именно как целевые классы для глубинной модели.

Важным шагом вперед с точки зрения масштаба и сложности является датасет WikiChurches, предложенный Barz и Denzler. Он построен на связке Wikidata и Wikimedia Commons и включает 9485 изображений 9346 европейских церквей. Каждый объект снабжен меткой архитектурного стиля, полученной из Wikidata, а стили организованы в трехуровневую иерархию: 64 стиля верхнего уровня (Gothic, Romanesque, Baroque, Renaissance и др.), 45 подстилей второго уровня (Gothic Revival, Brick Gothic, English Gothic и т.п.) и 8 стилей третьего уровня. На основе этой структуры выделяются несколько подмножеств: WikiChurches-14 (14 наиболее распространенных стилей), WikiChurches-6 (6 крупнейших классов - Gothic, Romanesque, Baroque, Renaissance, Classical Revival, Modern Architecture) и WikiChurches-4/19 (четыре основных стиля и 19 подстилей для задач иерархической классификации). Отдельной особенностью датасета является наличие 631 bounding box'a характерных элементов (tracery, арочные окна, контрфорсы, пинакли и др.), размеченных экспертом по истории архитектуры для четырех базовых стилей. Авторы подчеркивают сильную несбалансированность классов (две трети изображений приходятся на Gothic и Romanesque) и наличие около 6,5% шумных меток, что делает WikiChurches реалистичным, но сложным бенчмарком.

Наряду со стилевыми наборами данных формируются датасеты, ориентированные на анализ фасада как структуры и его материалов. В контексте семантической сегментации фасадов важны наборы ECP-Monge и CMP, содержащие выпрямленные изображения фасадов европейских зданий с пиксельной разметкой классов «окно», «дверь», «балкон», «стена», «крыша» и т.д. Эти наборы используются для оценки качества сегментации в работах по семантическому разбору фасадов.

В области анализа материалов фасадов Xie и соавт. представляют датасет BFMS (Building Facade Material Segmentation). Он включает 530 фотографий фасадов (в основном исторической застройки Китая), снятых на уровне человеческого взгляда, с детальной пиксельной разметкой 42 классов материалов и элементов окружения. Из них 26 относятся к собственно фасадным материалам и покрытиям (кирпич, камень, дерево, металл, стекло, штукатурка, черепица, гидроизоляционные материалы и др.), 16 - к категориям окружения (небо, вода, растительность, человеческая кожа и пр.). Средняя плотность разметки около 654 сегмента на изображение делает BFMS одним из наиболее детализированных фасадных датасетов по материалам.

Для анализа авторских стилей по работам отдельных архитекторов Zhong и соавт. создают датасет ArchDiffBench, включающий 1765 изображений, сгруппированных по десяти архитекторам (Мис ван дер Роэ, Ричард Майер, Ле Корбюзье, Ван Шу и др.), и

рассматривают 81 пару групп для попарного сравнения стиливых почерков. В отличие от предыдущих датасетов, здесь целевой сущностью выступает не исторический стиль, а почерк автора, однако методологически это близко к задачам классификации и анализа стилей.

Наконец, в работах Li и Dong используются данные не только в виде изображений, но и в виде CAD-моделей и BIM-описаний, содержащих информацию о геометрии, размерах, материалах и структурной схеме зданий. Хотя полноценного открытого датасета в статье не публикуется, она задает направление интеграции визуальных и параметрических данных.

2.3. Классические part-based и вероятностные модели стиля.

Одной из ключевых ранних работ по автоматической классификации архитектурных стилей является статья Xu и соавт., в которой на основе описанного выше 25-классового датасета вводится модель Multinomial Latent Logistic Regression (MLLR). Авторы используют деформируемые модели частей (Deformable Part-based Models, DPM) на HOG-признаках для представления фасада как набора частей: root-фильтра (силуэт фасада) и набора part-фильтров (окна, арки, купола, фронтоны и пр.), с учетом штрафов за деформацию частей относительно корня. В отличие от классического latent SVM, где для каждой модели обучается бинарный классификатор и на выходе получается жёсткое решение, MLLR формулируется как многоклассовая логистическая регрессия по латентным признакам.

Пусть $f(x, z)$ - это вектор признаков, зависящий от изображения x и латентной конфигурации частей z , а β_k - параметр класса k . Тогда для каждого класса определяется score-функция $(x; \beta_k) = \max_{z \in Z(x)} \beta_k^T f(x, z)$, а вероятности классов задаются стандартной softmax-функцией: $pk(x; \theta) = \frac{\exp(s(x; \beta_k))}{\sum_l \exp(s(x; \beta_l))}$. Обучение осуществляется максимумом регуляризованного правдоподобия с L1-регуляризацией (lasso), с использованием подхода по аналогии с latent SVM: по очереди фиксируются латенты положительных примеров и обновляются параметры, а затем пересчитываются латентные конфигурации негативов. Такая постановка позволяет, с одной стороны, учитывать латентную структуру фасада, а с другой - получать вероятностную оценку принадлежности изображения ко всем стилям одновременно.

В эксперименте со сжатыми десятью стилями, хорошо различимыми визуально, MLLR на основе DPM достигла точности порядка 67,8% и превзошла как классический DPM-LSVM (65,7%), так и глобальные методы (GIST, Spatial Pyramid, Object Bank, дававшие 30–63%). При расширении задачи до всех 25 стилей лучшим оказался гибридный метод MLLR+SP, объединяющий вероятностные выходы MLLR и глобальные признаки Spatial Pyramid; совокупная точность составляла около 46,2% и была выше, чем у DPM-LSVM (37,7%) и у отдельных глобальных признаков (44-45%). Кроме того, вероятностный характер MLLR позволил построить граф межстилевых отношений, показывающий, какие стили часто путаются и какие связаны исторически (following, revival, opposition), а также анализировать отдельные здания как смеси несколькими стиливыми влияний.

Дальнейшее развитие той же линии демонстрирует работа Zhao и соавт., предлагающая сложный гибридный конвейер DPM + глубокие признаки + модуль переразметки признаков. Авторы используют тот же 25-классовый датасет Хи, но существенно усиливают признаковое описание. На первом этапе изображения обрабатываются DPM: из исходного кадра выбирается наиболее репрезентативный фрагмент фасада, соответствующий маске максимального отклика DPM-модели (типичный купол, шпиль, фронтон, карниз). Это позволяет удалить небо, людей и другие нерелевантные области, концентрируясь на стиле.

Далее этот фрагмент подаётся в GoogLeNet, предобученную на ImageNet, которая извлекает сверточные признаки. Эксперименты показывают, что добавление глубинных признаков дает скачок точности примерно на 40 процентных пунктов по сравнению с использованием только DPM-наблюдений. Наконец, поверх глубинных признаков работает модуль Improved Ensemble Projection (IEP), который на основе предположений локальной согласованности и экзотической несогласованности формирует ансамбль прототипов - компактных групп изображений, представляющих отдельные визуальные концепции. Для этого применяется Max-Min sampling: сначала выбираются максимально разнесенные в пространстве признаков seed-образы (max-шаг), затем вокруг каждого seed формируется компактная группа ближайших соседей (min-шаг). Повторение этого процесса множество раз даёт богатое пространство атрибутов, по сходству с которыми формируется новое признаковое представление для каждого изображения. На этом представлении обучается линейный SVM.

Комплекс предобработка DPM + GoogLeNet + IEP + SVM демонстрирует на 25-классовом наборе Хи заявленную авторами точность порядка 98,6% (при 1000 обучающих изображений на класс), существенно превосходя все предыдущие DPM- и MLLR-основанные подходы (на уровне 40–55%). Хотя столь высокий результат потенциально связан с возможным переобучением, метод Zhao важен как пример гибридного решения, объединяющего структурную информацию о фасаде (DPM), глубокие признаки и специально обучаемое признаковое пространство, максимально сжимающее внутриклассовую вариативность.

2.4. Глубинные модели классификации стиля по глобальному изображению.

Переход к глубоким моделям без явной part-based структуры демонстрируют работы, в которых стили классифицируются по целому изображению здания или городского фрагмента.

Уже упомянутая работа Montoya Obeso и соавт. посвящена трехклассовой классификации мексиканских архитектурных стилей (prehispanic, colonial, modern) плюс класс «other». Авторы строят сравнительно неглубокую сверточную сеть (четыре сверточных и два полносвязных слоя) и исследуют влияние дополнительных геометрических признаков.

На базе датасета из 8000 исходных изображений и их аугментаций лучший результат (88,01% точности) достигается при использовании именно этих sparse features в комбинации с RGB и предложенной архитектурой; для сравнения, та же сеть на чистом RGB даёт 86,10%, AlexNet на RGB - около 74%, а добавление PCA-SIFT вместо sparse features приводит к падению точности до 65%. Таким образом, работа демонстрирует, что введение геометрических характеристик (углы, линии, их пересечения) может улучшать работу глубинной сети в задаче стиливой классификации, особенно при ограниченном объеме данных.

В более общем плане задачи классификации фасадов по стилю и возрасту рассматриваются в работах вроде «Understanding architecture age and style through deep learning». В таких исследованиях формируются сборные датасеты фасадов из разных городов (десятки тысяч изображений), а стили либо задаются в терминах исторических периодов строительства, либо отражают крупные стилевые категории. Обычно используются архитектуры семейства ResNet или Inception, иногда в мультитасковой постановке (одновременное предсказание возраста и стиля), а также CAM/Grad-CAM для визуализации важных участков фасада. Демонстрируется, что даже при шуме разметки и смещении стилей сети способны извлекать устойчивые паттерны и давать разумное качество на уровне десятков процентов точности при большом числе эпох/классов; однако проблема интерпретируемости и связи с архитектурной теорией остается открытой.

Работа Barz и Denzler с датасетом WikiChurches, хотя в основном и посвящена описанию самого набора данных, также включает базовые эксперименты по классификации стилей. Используются предобученные ResNet-50 и TResNet-M, дообучаемые на подмножестве WikiChurches-6. В качестве основной метрики берётся balanced accuracy (усредненная по классам точность), что важно в условиях сильной несбалансированности. Простое дообучение ResNet-50 на изображениях 224x224 даёт точность около 60%, undersampling по классам добавляет порядка 1,5 процентных пункта, а увеличение разрешения до 448x448, позволяющее лучше захватывать мелкие детали фасадов, поднимает показатель до 65,9%. Лучший результат (около 69,5% balanced accuracy) достигается при использовании архитектуры TResNet-M, предобученной на ImageNet-21k с иерархической многолейбловой схемой, и изображений 448x448. Даже при шести классах и мощной предобученной модели качество далеко от идеального, что подчеркивает сложность fine-grained классификации архитектурных стилей.

2.5. Семантическая сегментация фасадов и структурные признаки.

Понимание архитектурного стиля лишь по глобальному виду фасада ограничено: многие стилевые признаки зашиты в структуре фасадной композиции - ритме окон, расположении дверей, этажности, характере балконов, типах карнизов. Поэтому в ряде работ основное внимание уделяется семантическому разбору фасада на элементы.

Fathalla и Vogiatzis предлагают глубокий пайплайн для семантической сегментации фасадов, объединяющий полносверточную сеть VGG-16 и ограниченные Больцмановские машины (RBM), моделирующие архитектурный контекст по строкам и столбцам изображения. На первом этапе АСТ дает для каждого пикселя вероятности принадлежности к классам (стена, окно, дверь, балкон, крыша, небо). На втором этапе из этой разметки формируются скан-линии: последовательности пикселей по горизонтали и вертикали, кодируемые one-hot по классам. Эти скан-линии служат видимым слоем для двух RBM (горизонтальной и вертикальной), которые учатся распределению вероятных конфигураций фасадов (количество этажей, ритм окон, положение дверей и крыши). Затем формируется признак, по которому небольшой полносвязный MLP выдает уточненную сегментацию. Дополнительно вводится этап дообучения RBM на самом тестовом изображении. Эксперименты на датасетах ECP-Monge и CMP показывают, что добавление структурного слоя дает существенное улучшение по сравнению с одной только FCN: на ECP точность возрастает с 86,4% до 91,31%, на CMP - с 61,46% до 69,02%. В сравнении с другими на тот момент методами,

предложенный pipeline демонстрирует высокое качество, не опираясь при этом на жестко прописанные архитектурные правила (если балкон, то над ним окно), а извлекая их из данных.

Хотя эта работа не решает напрямую задачу классификации стилей, она предоставляет важный инструмент: возможность автоматически получать детализированную семантическую карту фасада, которая может служить базой для извлечения стилизованных признаков высокого уровня (отношение площади остекления к стене, наличие/отсутствие балконов определённого типа).

2.6. Анализ фасадных материалов и цвета.

Другую группу признаков стиля составляют материалы и цветовые решения. Они часто закреплены нормативами охраны наследия и воспринимаются как ключевые компоненты исторического облика. Работа Xie и соавт. посвящена разработке конвейера для автоматического анализа материалов и цветов фасадов исторической застройки и созданию датасета BFMS.

Модель BFMS основана на архитектуре Mask2Former со Swin Transformer в роли энкодера. Входные изображения нарезаются на патчи 768x768; далее пиксельный энкодер на Swin последовательно формирует пирамиду признаков, которые затем агрегируются в пиксельном декодере, а трансформер-декодер Mask2Former генерирует набор масок и соответствующих им классов материалов. Перед настройкой на BFMS модель предобучается на универсальном датасете материалов DMS, после чего 200 эпох доводится на BFMS.

На валидации BFMS достигает средней интерсекции по объединению (mIoU) около 0,415 и пиксельной точности порядка 76,5%. Для таких классов, как небо, кирпич, стекло, утрамбованная земля, кровельная гидроизоляция, IoU превышает 0,7–0,9, в то время как редкие и визуально неоднородные категории (мутный пластик, резина, керамика, кожа) распознаются хуже. Сравнение с Mask2Former, обученным только на DMS, показывает улучшение по большинству фасадных материалов, что подтверждает пользу специализированного фасадного датасета. По сравнению с архитектурами U-Net, DeepLabV3+ и SegFormer-B3 BFMS превосходит их по всем основным метрикам.

Помимо сегментации материалов, Xie и соавт. реализуют два важных модуля: извлечение текстур и извлечение цветов. Для текстур сегментированное изображение разбивается на сетку квадратов 100x100 пикселей; для каждого квадрата, где более 50% площади занято одним материалом, формируется кандидат на образцовую текстуру класса. Такие патчи сортируются и формируют библиотеку типичных фасадных текстур. Для цветов сначала выполняется коррекция баланса белого и экспозиции, затем пиксели каждого материала переводятся в HSV-пространство, где с помощью K-means (с ослабленным влиянием тёмных пикселей) извлекаются доминирующие цвета. Эти кластеры далее агрегируются по всем зданиям района для построения стандартной палитры материала и района.

Применение методики к историческому кварталу Taiping Alley (Цзиндэчжэнь, Китай) показывает, как полученные данные могут использоваться для количественной оценки стилистической согласованности застройки. По долям материалов на фасадах определяется,

какие здания соответствуют историческому образу (преобладание кирпича, камня, дерева, утрамбованной земли) и какие выбиваются (имитационный камень, керамическая плитка в качестве доминирующего материала). Сравнение с экспертными оценками показывает совпадение в подавляющем большинстве случаев. На основе анализа материалов и цвета формулируются обновлённые регламенты: допустимые материалы фасадов и допустимая цветовая гамма. Пример реализации - проект общественного центра, выполненный с учетом этих рекомендаций и органично вписывающийся в существующий исторический ландшафт.

В контексте классификации архитектурных стилей BFMS и аналогичные подходы не дают непосредственно ярлык стиля, но создают богатое признаковое пространство, отражающее материально-цветовую составляющую стиля. Совмещение этих признаков с геометрическими и структурными характеристиками фасада представляется перспективным направлением для более точной стилевой классификации.

2.7. Интеграция глубинных моделей и CAD/BIM.

Отдельную линию развития представляют работы, интегрирующие визуальные данные с CAD и BIM-моделями. Li и Dong рассматривают задачу классификации архитектурных стилей на основе данных архитектурного проектирования, подчеркивая, что стиль - это результат синтеза формы, конструктивной схемы, материалов и цвета, и что CAD-модели содержат часть этой информации в явном виде.

Предлагаемая ими модель включает сверточный энкодер, извлекающий визуальные признаки из изображений (фото или рендеров из CAD), модуль мультиуровневого слияния признаков, учитывающий как масштабные и структурные характеристики, так и параметры CAD (геометрия, размеры, материалы), и классификатор на основе опорных векторов. Обучение осуществляется с использованием стохастического градиентного спуска; подбор гиперпараметров (скорость обучения, размер батча) производится по результатам валидации (оптимальные значения — learning rate 0,01, batch size 256). На финальном этапе сверточный энкодер служит извлекательным модулем признаков, а SVM - классификатором.

Сравнение с классической моделью SVM (на ручных признаках) и моделью на основе рекуррентной сети показывает преимущество предложенного подхода: точность классификации превышает 97%, среднее абсолютное отклонение (MAE) около 1.24, а время обучения меньше, чем у обоих сравниваемых методов. В прикладном кейсе по оценке и охране древних зданий модель демонстрирует более 95% согласованности с экспертными оценками стиля.

Хотя работа не детализирует используемый датасет и перечень стилевых классов, она важна как пример интеграции визуальных и параметрических архитектурных данных и подчеркивает значимость мультиуровневых признаков (форма, структура, материалы) при классификации стилей.

2.8. Vision-language модели и интерпретируемый анализ стилового почерка.

Последний на сегодняшний день этап эволюции методов - применение больших vision-language моделей (VLM/VLLM) для анализа архитектурных стилей. Эти модели

объединяют визуальный и текстовый каналы, что позволяет переходить от немого ярлыка стиля к объяснимым текстовым описаниям стилиевых особенностей.

Zhong и соавт. предлагают рамку ArchiLense для количественного анализа авторских стилей на базе датасета ArchDiffBench. Фреймворк состоит из двух подсистем: Style Extractor и Style Evaluator. Первая автоматически генерирует текстовые описания различий между двумя группами изображений (проектами двух архитекторов), вторая количественно проверяет, насколько эти описания действительно различают группы.

В Style Extractor рассматриваются несколько вариантов, основанных на VLM: прямой анализ изображений, генерация разностей эмбедингов и последующее текстовое описание, а также вариант, признанный наиболее успешным - сопоставление наборов текстовых подписей, автоматически сгенерированных для обеих групп, и генерация на их основе описания стилистических отличий. В Style Evaluator для каждого текстового описания у вводится мера соответствия $\theta(x, y)$ между изображением x и текстом (на практике - косинусное сходство между эмбедингами), и на её основе вычисляется разность средних по двум группам. По распределениям $\theta(x, y)$ для групп А и В проводится t-тест; описания, для которых различие статистически значимо ($p < 0.05$), признаются дифференцирующими.

Экспериментально более 80% описаний, сгенерированных для 270 пар, оказываются статистически значимыми.

2.9. Итог.

С учетом вышеописанного, ряд исследовательских вопросов остаётся открытым. Во-первых, требуется интеграция различных уровней признаков - глобальной геометрии, структурной семантики фасада (окна, двери, балконы, карнизы), материалов и цвета в единую модель классификации архитектурных стилей. Во-вторых, мало исследованы иерархические постановки, учитывающие как широкие эпохи, так и подстили, а также возможность смешения нескольких стилиевых влияний в одном здании. В-третьих, на фоне успехов VLM остается во многом нерешенной задача сочетания высокой точности классификации с интерпретируемостью, когда модель не только выдает ярлык стиля, но и объясняет его через понятные архитектору структурные, материальные и цветовые признаки.

Данная работа, посвященная классификации архитектурных стилей, может занять свое место в этом поле, опираясь на наработки по семантической сегментации фасадов и анализу материалов (как источник содержательных признаков), на классические и гибридные модели стилиевой классификации (как отправную точку для выбора архитектуры) и на идеи интерпретируемого анализа (как ориентир при формулировании выходных характеристик модели). Такой синтез позволит не только улучшить качество автоматической классификации стилей, но и связать результаты работы с традиционной архитектурной теорией.