

Обзор методов спектральной реконструкции по трёхканальным изображениям¹

Д. М. Владимиров^{*,**}, Д. А. Реутский^{*,**,***}, Е. И. Ершов^{*,**}

^{*}Институт проблем передачи информации имени А.А. Харкевича, Российская академия наук,
Москва, Россия

^{**}Московский физико-технический институт (национальный исследовательский университет),
Москва, Россия

^{***}Вюрцбургский университет, Вюрцбург, Германия

01.10.2025 г. Принята 30.10.2025 г.

Аннотация—Гиперспектральные изображения (ГСИ) хранят в каждом своем пикселе распределение интенсивности света по длинам волн. В связи с высокой стоимостью оборудования для съемки ГСИ, исследователи разрабатывают методы оценки спектральных характеристик объектов в сценах по сжатым или маломерным представлениям спектров. Такой процесс оценки принято называть гиперспектральной реконструкцией (ГР). В литературе часто рассматривается возможность оценить ГСИ по доступным RGB-изображениям сцены. В настоящей работе представлен обзор и классификация самых значимых из этих методов, а также изложены общепринятые наборы данных и метрики для оценки качества реконструкции из RGB-изображений. Проведена вычислительная ревизия: все методы заново обучены и протестированы на наборе данных ARAD 1K. В литературе нет общепринятой постановки задачи ГР, не уточняется природа трёхканальных изображений, должны они быть в RAW-формате или JPEG, эксперименты выполнены для обоих вариантов. Наилучшие достигнутые значения PSNR 32,0 dB и 33,5 dB при реконструкции из JPEG-изображений и RAW-изображений соответственно. При смене тестового набора данных на CAVE у всех методов наблюдается ухудшение качества реконструкции минимум в 2 раза: это показывает, что низкое разнообразие сцен в ARAD 1K ограничивает обобщающую способность методов ГР. Также в работе обсуждаются возможные направления будущих исследований и текущие проблемы, возникающие при решении поставленной задачи.

КЛЮЧЕВЫЕ СЛОВА: гиперспектральная реконструкция, спектр, нейронные сети, трансформеры

DOI: 10.53921/18195822_2025_25_3_427

1. ВВЕДЕНИЕ

Гиперспектральные изображения (ГСИ) содержат богатую спектральную информацию об объектах в сцене: в каждом их пикселе зафиксирована зависимость интенсивности наблюдаемого излучения от длины волны. Такое представление позволяет анализировать физико-химические свойства объектов в сцене и находит широкое применение в различных областях: медицине [1], дистанционном зондировании [2, 3], исследовании исторических документов [4], пищевой промышленности [5, 6], а также при разработке алгоритмов обработки RAW-изображений [7, 8].

Для получения точных ГСИ обычно используются специализированные гиперспектральные камеры. Многие из них сканируют сцену «построчно», пропуская по очереди её узкие

¹ Работа выполнена в рамках государственного задания №FFNU-2025-0045 ИППИ РАН, утвержденного Минобрнауки России.

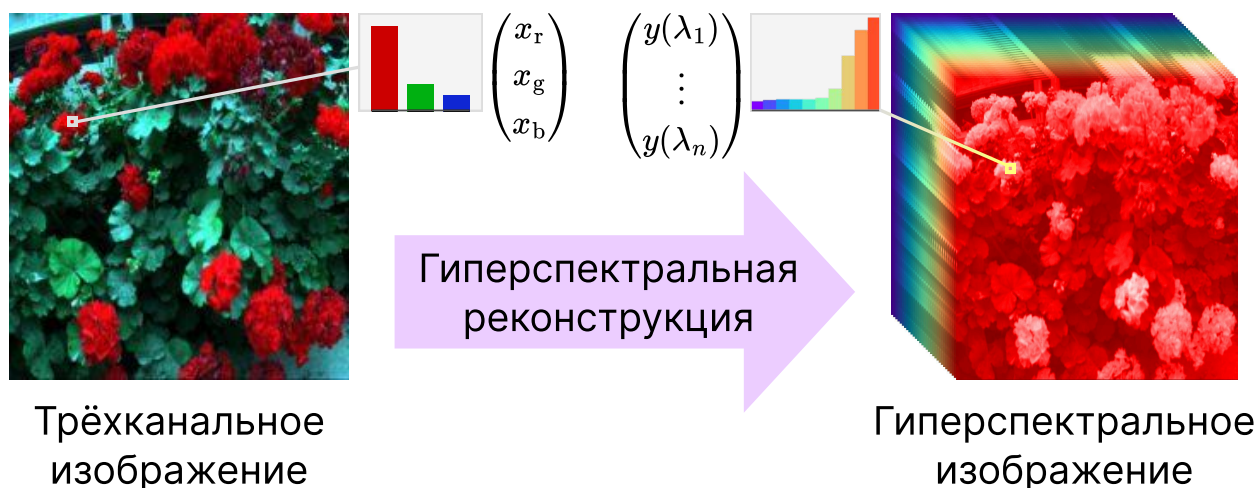


Рис. 1. Реконструкция гиперспектрального изображения из трёхканального.

фрагменты через дифракционную решётку [9, 10]. Распространены также акустооптические камеры [11, 12, 13], параметры дифракционной решетки в которых настраивается за счет стоячей волны в кристалле. Несмотря на качество получаемых данных, такие камеры труднодоступны по причине высокой стоимости, а также имеют длительное (вплоть до нескольких минут) время съемки — оба эти фактора ограничивают потенциал их распространения для решения практических задач.

Альтернативным способом получения ГСИ может быть их оценка из некоторого маломерного или закодированного представления, известная в литературе как гиперспектральная реконструкция (от англ. hyperspectral reconstruction). Например, ГСИ высокого пространственного разрешения можно восстановить, используя пару изображений: одно с высоким пространственным, но низким спектральным и одно ГСИ с низким пространственным разрешением и высоким спектральным [14, 15]. Другой пример — съёмка кодирующей апертурой (от англ. Coded Aperture Snapshot Spectral Imaging, CASSI), которая затеняет отдельные мелкие участки сцены, после чего свет сцены проходит через дисперсный элемент [16, 17, 18]. Каждый спектр превращается в вытянутую радужную полоску, а благодаря затенению не все они накладываются друг на друга. В результате на монохромной фотоматрице получается изображение, на котором спектральная информация закодирована вдоль определённой пространственной оси.

Оценку ГСИ также можно проводить из изображений с низким спектральным разрешением и воспринимать эту задачу, как интерполяцию спектральной информации в каждом пикселе. Доступным вариантом изображений со сжатой спектральной информацией являются RGB-изображения, съёмку которых может произвести каждый владелец современного смартфона. Далее в этой работе гиперспектральной реконструкцией (ГР) будем называть задачу оценки ГСИ по RGB-изображениям [19, 20], если не сказано обратного. Обычно спектр восстанавливают в видимом диапазоне, примерно 400–700 нм [20], так как на линзах цифровых камерах стоят оптические фильтры, поглощающие свет в ультрафиолетовом (УВ) и инфракрасном (ИК) диапазонах. Задача ГР из RGB-изображений проиллюстрирована на Рис. 1.

С математической точки зрения задача ГР является некорректной обратной задачей, подробное описание и постановка задачи даны в разделе 2. Некорректной она является из-за отсутствия единственного решения. С физической точки зрения множественность решений вызвана таким явлением, как *метамеризм*: для каждого цветового отклика существует бесконечное множество дающих его спектров — метамеров. Тем не менее, есть предпосылки, позволяющие оценивать спектр по цветовому отклику с неплохой точностью. Фактически ре-

ализуемые в природе спектры имеют низкую внутреннюю размерность или, иными словами, лежат на маломерном многообразии. Было показано, что примерно 7-8 главных компонент PCA или 3 компоненты Isomap описывают 99% дисперсии в наборах данных гиперспектральных изображений [21, 22]. Другое исследование подсказывает что, среди всех пар спектров из одного ГСИ только 0,01% будут метамерами [23].

Для решения задачи ГР было предложено множество различных подходов — от классических модельных методов, основанных на предположениях о структуре многообразия спектров [22, 24, 25, 26, 21, 27], до современных нейросетевых решений [28, 29, 30, 31, 32, 19, 33, 34, 35, 36, 18, 37]. Основным драйвером развития индустрии являются соревнования, проводящиеся в рамках международных конференций [38, 39, 20, 40]. Именно для этих соревнований собраны самые большие наборы ГСИ повседневных объектов и предложены state-of-the-art (SotA) решения.

В данной работе представляется обзор существующих постановок задачи ГР, методов ее решения и соответствующих наборов данных. Близкая по тематике работа была представлена в 2022 году [41]. Настоящая работа дополняет её, включая рассмотрение нейронных сетей на базе трансформерной архитектуры, предложенных за последние три года, а также предоставляя унифицированную кодовую базу для воспроизводимости результатов и разработки новых методов. Предлагается также расширить классификацию методов ГР, сделав основной акцент на способе оценки спектральных характеристик сцены и на способе обработки признаков. Кроме того, обсуждаются ограничения существующих подходов, главное из которых заключается в маленьком размере существующих наборов данных и однотипности. Что в купе со способностью нейронных сетей детектировать шумовой профиль сенсора приводит к переобучению под существующие наборы данных, и препятствует генерализации разработанных решений. Также обсуждаются возможные направления дальнейших исследований и способы борьбы с текущими проблемами.

Код доступен на github по ссылке: <https://github.com/createcolor/HSRReview>.

2. ПОСТАНОВКА ЗАДАЧИ СПЕКТРАЛЬНОЙ РЕКОНСТРУКЦИИ

Естественно описывать многоканальное изображение трёхмерным массивом $\mathbf{X} \in \mathbb{R}^{h \times w \times m}$, где h, w — высота и ширина изображения соответственно. Для гиперспектрального изображения используется схожее описание массивом $\mathbf{Y} \in \mathbb{R}^{h \times w \times n}$: в пикселе с координатами (i, j) хранится вектор $\mathbf{y}_{ij} \in \mathbb{R}^n$, являющийся упрощённым дискретным представлением исходного спектра-стимула $y_{ij}(\lambda)$, освещавшего фотопиксель (i, j) в момент съёмки. Под дискретностью понимается выбор конечного числа длин волн $\lambda_1 < \dots < \lambda_n$ и замена всей функции $y_{ij}(\lambda)$ на её значения в каждой λ_k или на интегралы функции в окрестности каждой точки. Обычно в качестве решетки выбирается 31 длина волны от 400 до 700 нм с шагом 10 нм. Таким образом, $\mathbf{y}_{ij} = [y_{ij}(\lambda_k)]_{k=1}^n$. Далее в тексте аналогично будет обозначаться за $\mathbf{x}_{ij}$ пиксель с координатами (i, j) у многоканального изображения, а также иногда будут опускаться индексы ij , если контекст позволяет обойтись без них. Вообще выбор m ничем неограничен и может принимать любые натуральные значения, важно лишь, что $m \ll n$. В литературе самая распространенная постановка — реконструкция из RGB-изображений. Поэтому далее в работе будем полагать $m = 3$.

Задача гиперспектральной реконструкции — получить оценку $\hat{\mathbf{Y}}$ по $\mathbf{X}$. Для оценки качества реконструкции, а также для обучения методов с учителем требуется истинное значение $\mathbf{Y}$, попиксельно сопоставленное с $\mathbf{X}$. Получить такую пару $(\mathbf{X}, \mathbf{Y})$ в реальных условиях затруднительно, поэтому исследователи обходятся лишь съёмкой $\mathbf{Y}$ на гиперспектральную камеру, а затем синтетической генерацией $\mathbf{X}$ из $\mathbf{Y}$. Таким образом, все количественные оценки ГР получены на синтетических входных данных. Оценки методов ГР на реальных входных данных

тоже существуют, но производятся менее тщательно: из $\mathbf{X}$, снятого на реальную трехканальную камеру, реконструируется $\hat{\mathbf{Y}}$, из которого выбирается пиксель $\hat{\mathbf{y}}$, а на $\mathbf{Y}$ ищется визуально соответствующий ему $\mathbf{y}$, после чего $\hat{\mathbf{y}}$ и $\mathbf{y}$ сравниваются визуально и количественно.

Для описания модели формирования изображения рассматривается оптическая система камеры, фокусирующая на пикселе фотоматрицы видимое излучение со спектром $y_{ij}(\lambda)$. Он формируется из суммарного спектра освещения источником света $l(\lambda)$, а также спектральной отражательной способности участка объекта сцены $r_{ij}(\lambda)$, соответствующего будущему пикселю изображения: $y_{ij}(\lambda) = l(\lambda) \cdot r_{ij}(\lambda)$. Здесь делается предположение о том, что все поверхности Ламбертовы, то есть рассеивают свет одинаково во всех направлениях.

Обозначим спектральную чувствительность фотопикселя за $c_k(\lambda)$, где k — его тип: красный, зелёный или синий. Тогда полная эффективная мощность излучения, воздействующего на фотоэлемент пикселя, даётся выражением

$$x_k = \int_{\Lambda} c_k(\lambda) l(\lambda) r(\lambda) d\lambda, \quad k = r, g, b, \quad (1)$$

где Λ — диапазон длин волн, видимых данным фотосенсором, вне него $\forall k \ c_k(\lambda) = 0$. Чаще всего в задаче ГР происхождение $y(\lambda)$ не имеет значения. Тогда (1) можно переписать в виде

$$x_k = \int_{\Lambda} c_k(\lambda) y(\lambda) d\lambda. \quad (2)$$

Дискретизируя $c_k(\lambda)$ по аналогии с $y(\lambda)$, рассматривается приближенная модель формирования сигнала в фотопикселе

$$x_k = \mathbf{c}_k^T \mathbf{y}. \quad (3)$$

В (3) содержатся три уравнения для всех k . Собирая из $\mathbf{c}_k$ матрицу C , модель формирования изображения записывается в матричном виде

$$\mathbf{x} = C\mathbf{y}. \quad (4)$$

Синтетическое трёхканальное изображение $\mathbf{X}$ получается применением (4) к каждому пикселю $\mathbf{y}$ гиперспектрального изображения $\mathbf{Y}$. В данной работе полученное $\mathbf{X}$ иногда будет называться *RGB-проекцией* $\mathbf{Y}$, хотя оператор C не является проектором в терминологии линейной алгебры.

Для приближения чистого симулированного сигнала $\mathbf{X}$ к цветовым откликам, возникающим в реальном мире, к такому чистому сигналу применяют дополнительные операции искажающие входной сигнал. Разные исследователи далее выборочно применяют подмножество из нижеуказанных операций, для упрощения записи у x_{ijk} и $\mathbf{x}_{ij}$ иногда опущены индексы:

1. Добавление шума. Распространена модель в виде суммы пуассоновской и гауссовской компонент [42, 20, 43].
2. Переход в цветовое пространство linear sRGB: $\mathbf{x}' = M\mathbf{x}$, где M — матрица перехода [37].
3. Нормализация на максимум: $\mathbf{X}' = \mathbf{X}/(\max_{ijk} x_{ijk})$ [19].
4. Гамма-коррекция: $x' = x^\gamma$, $\gamma = 0.416$ [37].
5. Приведение к 8-битности: $x' = \lfloor 255x \rfloor / 255$ [20].
6. Сжатие: $\mathbf{X}' = \text{JPEG}(\mathbf{X})$ [20].
7. Хроматические aberrации — свертка со спектральной функцией рассеяния точки (от англ. point spread function, PSF) при симуляции RGB-откликов [44].

Выбор матрицы C тоже остаётся на усмотрение исследователя, а потому отличается от работы к работе.

3. ТАКСОНОМИЯ СОВРЕМЕННЫХ МЕТОДОВ ГИПЕРСПЕКТРАЛЬНОЙ РЕКОНСТРУКЦИИ

В данном разделе описаны принципы работы избранных методов ГР из RGB-изображений. Верхнеуровнево все методы можно классифицировать по двум группам:

1. *Методы с априорно заданной структурой* используют некоторое модельное представление о спектрах. Например, задают множество, которому может принадлежать $\hat{\mathbf{y}}$. Как правило, работают с пикселями $\mathbf{x}$ входного изображения $\mathbf{X}$ независимо, но могут включать в его признаковое описание окружающий фрагмент 3×3 , 5×5 и т. д.
2. *Нейросетевые методы* являются параметризованными функциями $f_\theta : \mathbb{R}^{h \times w \times 3} \rightarrow \mathbb{R}^{h \times w \times n}$ наперёд заданного вида. Поскольку данные методы принимают обычно на вход всё изображение $\mathbf{X}$ целиком, они способны учитывать глобальный контекст. При дизайне архитектур нейронных сетей спектральная модель не задается явно, в некоторых случаях учитываются общие соображения о природных закономерностях спектров. Например, корреляция соседних компонент y_k и y_{k+1} мотивировала авторов MST++ [19] использовать спектральный механизм самовнимания.

Полная классификация методов ГР представлена на Рис. 2.

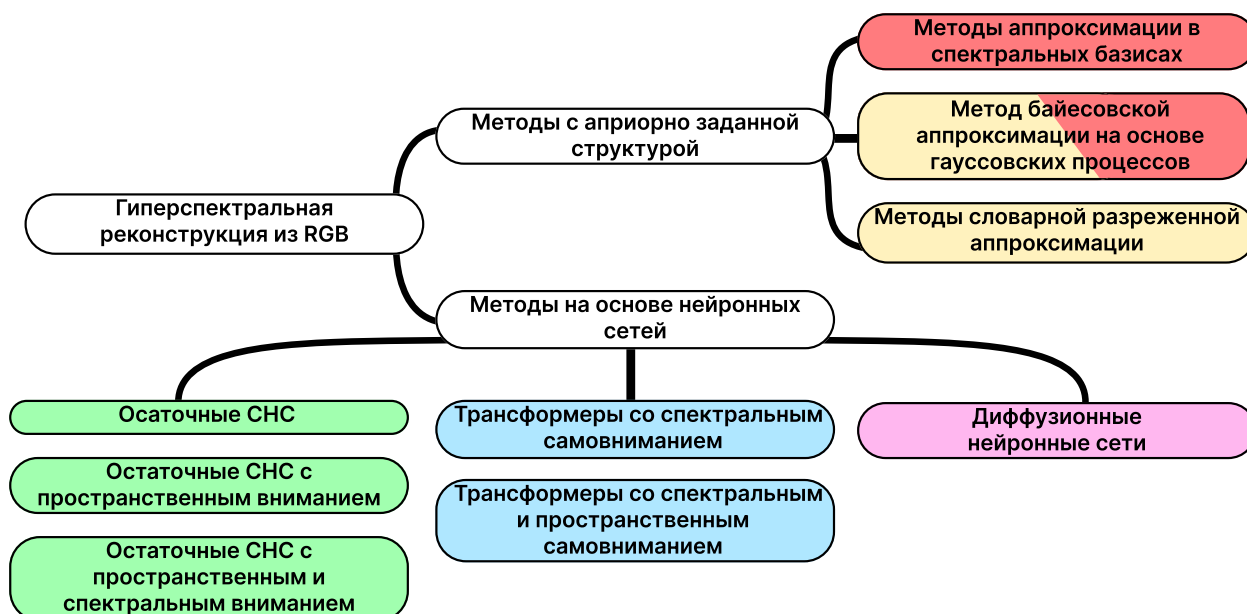


Рис. 2. Классификация методов ГР. Значение цветов на диаграмме: красный — методы ограничивающие класс функций спектров, реконструкция делается за счет оценки параметров модели по RGB-отклику; жёлтый — методы пользующиеся предположением о малой размерности внутреннего представления спектрального пространства, обучение заключается в выборе репрезентативного набора спектров, разреженной комбинацией которых можно приблизить все остальные; зелёный — методы на основе свёрточных нейронных сетей (СНС), то есть внутреннее представление изображений обрабатывается свёрточными блоками нейронной сети; синий — методы на основе трансформерных нейросетей; розовый — нейронные сети использующие механизм диффузии для обработки внутреннего представления изображений.

3.1. Методы с априорно заданной структурой

Словарная разреженная аппроксимация.

Ряд методов ГР основан на словарном обучении, одном из видов обучения представлений без учителя. В нём оптимизируется переполненный базисный набор — словарь, по которому любой элемент набора данных можно разложить в разреженную линейную комбинацию. Формально задача поиска такого словаря ставится в виде:

$$\min_{D, W} \|Y - DW\|_F \quad (5)$$

$$\text{s.t. } \forall j \|\mathbf{w}_j\|_0 \leq t, \quad (6)$$

где $Y \in \mathbb{R}^{n \times N}$ — набор данных из N элементов размерности n , $D \in \mathbb{R}^{n \times K}$ — искомый словарь размера K , $W \in \mathbb{R}^{K \times N}$ — матрица коэффициентов разложения по словарю, запись $\|\cdot\|_0$ означает число ненулевых компонент вектора, K, t — гиперпараметры. Для решения задачи (5) был разработан алгоритм K-SVD [45].

В контексте задачи ГР идея обучать словарь для разреженного представления используется в предположении, что каждый материал в сцене состоит из малого числа веществ. Для моделирования спектра такой смеси предлагается модель линейного смешивания. Соответственно, обучение словаря D — это попытка вычислить первичные спектры, из смеси которых состоят все остальные, а разреженность ограничивает число веществ, которое одновременно могут присутствовать в материале [46].

Arad — один из первых словарных методов ГР был предложен в 2016 году Арадом и Бен-Шахаром [22]. Словарь D_y обучается алгоритмом k-SVD на наборе тренировочных спектров. Затем формируется соответствующий словарь в пространстве цветов $D_x = CD_y$, аналогично формуле (4). На стадии реконструкции тестовый цветовой отклик $\mathbf{x}$ раскладывается в разреженную линейную комбинацию $\mathbf{x} = D_x \mathbf{w}$ методом ОМР [47]. Предсказанием алгоритма назначается $\hat{\mathbf{y}} := D_y \mathbf{w}$.

A+ был предложен Эшбахером и др. [24] в 2017 году. Первый шаг обучения такой же, как в Arad — построения D_y и D_x . Для каждого элемента D_x находятся с ближайших цветовых откликов из обучающего набора, и между ними и соответствующими спектрами обучается гребневая регрессия. Таким образом, получается k регрессий. На этапе тестирования $\mathbf{x}$ подаётся на вход регрессии, соответствующей ближайшему к $\mathbf{x}$ элементу D_x . Другими словами, метод выучивает многомерное кусочно-линейное отображение из цветов в спектры. Авторы также предлагают расширить $\mathbf{x}$ до 27-мерного признакового представления, добавив туда значения RGB окружающего фрагмента 3×3 , и демонстрируют прирост качества ГР от этого.

Обучение многообразия — подход Цзя и др. [21] подразумевает явное использование предположения о малой размерности многообразия натуральных спектров. Сперва авторы понижают размерность алгоритмом Isomap [48], получая трёхмерные представление спектров. Затем они учат сеть радиально-базисных функций (RBFN) отображению из RGB в обученное представление спектров. Недостающее звено — переход из данного представления обратно в спектры — реализовано словарным методом [49].

Гауссовские процессы.

Идея Ахтара и Миана [27] заключается в аппроксимации функций спектров-стимулов гауссовскими процессами для учета гладкости спектров, скоррелированности интенсивностей и шума в близких спектральных каналах. Тренировочная процедура метода начинается с разрезания всех тренировочных ГСИ на фрагменты 8×8 , которые кластеризуются методом k-means. Дальнейшие шаги тренировки происходят изолированно внутри каждого кластера. Данный приём призван уменьшить количество метамеров в обучающем наборе, так как спектры внутри кластера, скорее всего, будут очень схожи по происхождению, например, будут взяты из

травы, а значит, уменьшится размерность их многообразия. Обозначим центр кластера за $P \in \mathbb{R}^{n \times 64}$. На этапе тестирования нам понадобится его образ CP в пространстве RGB.

Из совокупности фрагментов внутри кластера случайно выбирается 1% пикселей, которые формируют обучающий набор спектров Y^* . Они рассматриваются как реализации случайного процесса

$$\mathbf{y} := \Phi(\mathbf{z} \odot \mathbf{s}) + \boldsymbol{\epsilon}, \quad (7)$$

где Φ — матрица, столбцами которой являются $\phi_i \sim \text{Norm}(\boldsymbol{\mu}_i, \Sigma_i)$, $z_j \sim \text{Bern}$, $s_j \sim \text{Norm}$, $\boldsymbol{\epsilon} \sim \text{Norm}$, Σ_i — квадратичное экспоненциальное ядро, $\odot$ — покомпонентное произведение. Для упрощения изложения указаны не все параметры распределений: полное описание вероятностной модели выходит за рамки данного обзора. Отметим только, что опущенные параметры тоже являются случайными величинами, чьи распределения зависят от других параметров.

В выражении (7) компоненты $\mathbf{z}$ принимают значения $\{0, 1\}$, что явно задаёт спектр в виде разреженной комбинации столбцов Φ и перекликается с идеей разреженного кодирования. Квадратичное экспоненциальное ядро обеспечивает гладкость спектров, наблюдаемую в реальном мире.

Байесовским выводом путём сэмплирования по Гиббсу авторы получают уточненные гиперпараметры $\{\boldsymbol{\mu}_i^*\}_{i=1}^K$ при условии наблюдения Y^* . В качестве начального предположения о значении $\boldsymbol{\mu}_i$ берутся столбцы матрицы D из решения задачи

$$\min_{D, W} \|Y^* - DW\|_F \quad (8)$$

$$\text{s.t. } \forall i, j \|\mathbf{w}_j\|_1 \leq \delta, \mathbf{d}_i, \mathbf{w}_j \geq 0, \quad (9)$$

которая решается алгоритмом из [50].

Таким образом, данный метод похож на метод разреженного кодирования, но с операцией уточнения байесовским выводом. Уточнённый словарь обозначим за $D^* := [\boldsymbol{\mu}_i^*]_{i=1}^K$.

На этапе тестирования входное RGB-изображение $\mathbf{X}$ рассматривается как набор фрагментов 8×8 , сдвигаемых на 2 пикселя по каждой координате. Алгоритм восстановления работает с каждым фрагментом независимо. Окончательное изображение $\hat{\mathbf{Y}}$ формируется усреднением перекрывающихся фрагментов.

Входному фрагменту назначается кластер, чей CP ближе всего по евклидову расстоянию. Соответствующий выбранному кластеру словарь D^* и пиксель из входного фрагмента $\mathbf{x}$ подставляются в задачу

$$\min_{\mathbf{w}} \|\mathbf{w}\|_1 \quad (10)$$

$$\text{s.t. } \|\mathbf{x} - CD^*\mathbf{w}\|_2^2 \leq \delta', \quad (11)$$

после решения которой предсказанием назначается $\hat{\mathbf{y}} := D^*\mathbf{w}$. Процедура повторяется для всех пикселей фрагмента.

Отметим, что идея с кластеризацией фрагментов для уменьшения метамеризма потенциально применима к любому методу ГР, работающему попиксельно.

Спектральные модели.

Существует подход описания спектров, разрабатываемый Николаевым [26] и Максимовым [51]. Вместо оценки спектра как произвольной функции интенсивности от длины волны, Николаев предлагает вводить параметрические модели спектров. Задача оценки спектра таким образом сводится к оценке его параметров. Николаев выделяет 10 требований к спектральным моделям [26]. Основные из них заключаются в том, что модель должна быть замкнута относительно операций, естественно возникающих в природе. Например, операция линейной

комбинации соответствует смешиванию двух и более источников света. Операция умножения описывает отражение света от различных поверхностей. Если спектральная модель определяется тремя параметрами, появляется возможность соотнести спектр трёхмерному цветовому отклику.

Важно заметить, что данные модели разрабатываются в контексте решения задачи цветовой константности (ЦК) — описания цветов объектов независимо от освещения — и не были изучены за пределами этой области.

3.2. Нейросетевые методы ГР

С развитием нейросетевых подходов задача ГР стала одной из задач перевода изображения в изображение (от англ. image-to-image), принцип решения которых состоит в обучении нейронной сети для перевода изображения из одного домена в другой путём минимизации функции потерь. При этом все предположения о структуре входных и выходных данных используются не для их предобработки, пример из предыдущей главы — построение словаря, а в архитектуре нейронной сети. В данном разделе описаны наиболее известные нейросетевые методы ГР. Все методы были обучены с использованием оптимизатора Adam [52], если не сказано обратного.

Свёрточные нейронные сети.

HSCNN-D [28] — первая глубокая свёрточная нейронная сеть для решения задачи ГР, использующая остаточные соединения и плотную структуру. Остаточные соединения были разработаны как решение проблемы затухающих градиентов. Плотная структура подразумевает, что каждый следующий блок нейронной сети принимает на вход конкатенацию признаков, выделенных всеми предшествующими блоками. Базовый свёрточный блок HSCNN-D для выделения признаков показан на Рис. 3. Нейронная сеть завершается свёрточным слоем с ядром размера 1×1 , выполняющим финальное преобразование карты признаков в ГСИ. Для обучения используется функция потерь L_2 между восстановленным и реальным ГСИ.

HRNet [29] использует пирамидальную структуру для выделения признаков из входных изображений. Для масштабирования изображения используются операции PixelUnshuffle и PixelShuffle, которые позволяют делать это без потери информации в отличие от свёрточного слоя с шагом или пулингового слоя (от англ. pooling layer). На каждом уровне обработка признаков происходит за счёт двух блоков — остаточного свёрточного блока, состоящего из пяти последовательных свёрточных слоев с ядром размера 3×3 и остаточного MLP-блока, который призван выделять глобальную карту внимания для признаков на текущем уровне нейронной сети. Для обучения авторы использовали функцию потерь L_1 .

AWAN [30] в качестве базового блока использует двойной блок остаточного внимания (от англ. Dual Residual Attention Block; см. Рис. 3). В отличие от предыдущих методов в AWAN используются два набора признаков: основной и остаточный. Основной набор признаков обрабатывается свёрточными блоками с остаточным соединением, а остаточный только прибавляется к основному на выходе свёрточных блоков, но не обрабатывается отдельно. Авторы утверждают, что такой подход позволяет сети не забывать низкочастотную информацию из входных изображений. В конце каждого базового блока применяется свёрточный слой с ядром размера 3×3 , после чего с помощью адаптивного взвешивания спектральных каналов выделяются спектральные карты внимания (Adaptive Weighted Channel Attention (AWCA) на Рис. 3. Отображением выделенных признаков в ГСИ занимается модуль пространственного внимания. Функция потерь состояла из двух слагаемых: MRAE между восстановленным и реальным ГСИ и L_1 между проекциями реального и восстановленного ГСИ в пространстве CIE XYZ. Определение MRAE дано в Разделе 5.1.

HDNet [31] представляет собой глубокую остаточную свёрточную сеть, в середине которой стоит модуль сопряжения спектрального и пространственного внимания. К входным признакам данного модуля применяются карты внимания вдоль пространственных и спектральной осей, полученные признаки суммируются и поканально обрабатываются остаточными свёрточными блоками. В качестве функции потерь авторы использовали сумму L_1 между восстановленным и реальным ГСИ и взвешенного L_1 между Фурье-образами данных изображений, где больший вес давался более высокой частоте.

Авторы **SSRNet** [32] ставят задачу ГР как минимизацию L_2 между восстановленным и реальным ГСИ с двумя регуляризационными множителями: пространственным и спектральным. Для решения такой задачи применяется метод множителей с переменным направлением (от англ. Alternating Direction Method of Multipliers, ADMM). Каждый шаг оптимизации предлагается аппроксимировать одним блоком нейронной сети. На i -ом шаге из приближения ГСИ выделяется два набора признаков — пространственные обычным свёрточным слоем и спектральные модулем AWCA, после чего происходит слияние данных признаков с использованием таких же модулей. Иллюстрация общей архитектуры сети показана на Рис. 3. В качестве функции потерь авторы использовали L_1 .

Трансформерные нейронные сети.

TSA-Net. Для решения задачи ГР в работе [18] Мэн и соавт. предлагают U-Net-архитектуру, декодеры которой используют трансформеры. На каждом уровне декодера с помощью свёрточных слоев выделяются признаки, по которым считается пространственное самовнимание. Параллельно с ним схожим образом подсчитывается спектральное самовнимание. Полученные признаки матрично перемножаются вдоль соответствующих осей. В качестве функции потерь использовались L_1 между восстановленным и реальным ГСИ и L_1 между производными спектров [53].

MST++. В работе [19] Цай и соавт. предложили использовать трансформерную архитектуру для решения задачи ГР. В отличие от классического Vision Transformer (ViT), который в качестве токенов использует фрагменты изображения, в MST++ используется механизм спектрального многоголового самовнимания (от англ. Spectral-wise Multi-head Self-Attention, S-MSA), который в качестве токена использует срезы изображения по каждому каналу. Такой подход позволил сохранить глобальное перцептивное поле нейронной сети и обеспечивает сложность вычисления механизма самовнимания $O(hw)$, в отличие от $O(h^2w^2)$ у ViT. Трёхуровневая U-Net-подобная архитектура MST++ изображена на Рис. 4. На каждом уровне применяется механизм спектрального самовнимания с последующим за ним Feed-Forward-блоком. Для выделения начальных признаков из входного RGB-изображения и отображения внутренних нейросетевых признаков используются свёртки 3×3 . В качестве функции потерь авторы использовали MRAE.

SPECAT. Чжиян и соавт. [33] используют подход, похожий на MST++, но на каждый уровень U-Net перед спектральным вниманием добавляют механизм иерархического пространственного внимания (от англ. Hierarchical Spatial Attention, HSA). HSA-блок является Swin-трансформером [54]. В качестве функции потерь авторы использовали сумму L_2 между восстановленным и реальным ГСИ и L_2 между их RGB-проекциями.

MSFN. Многоступенчатый трансформер MSFN, предложенный Бу и др. [34] так же, как и SPECAT, использует Swin-трансформер для извлечения и обработки пространственных признаков ГСИ и — по аналогии с MST++ — S-MSA для спектральных. MSFN состоит из трех одинаковых пар пространственных и спектральных U-Net-образных блоков. Чтобы не допустить затухания градиентов и потери информации из каждого уровня «пространственного» U-Net проходит остаточное соединение на такой же уровень «спектрального» U-Net. Так же, как и в MST++, для извлечения начальных признаков из входного изображения и проекции

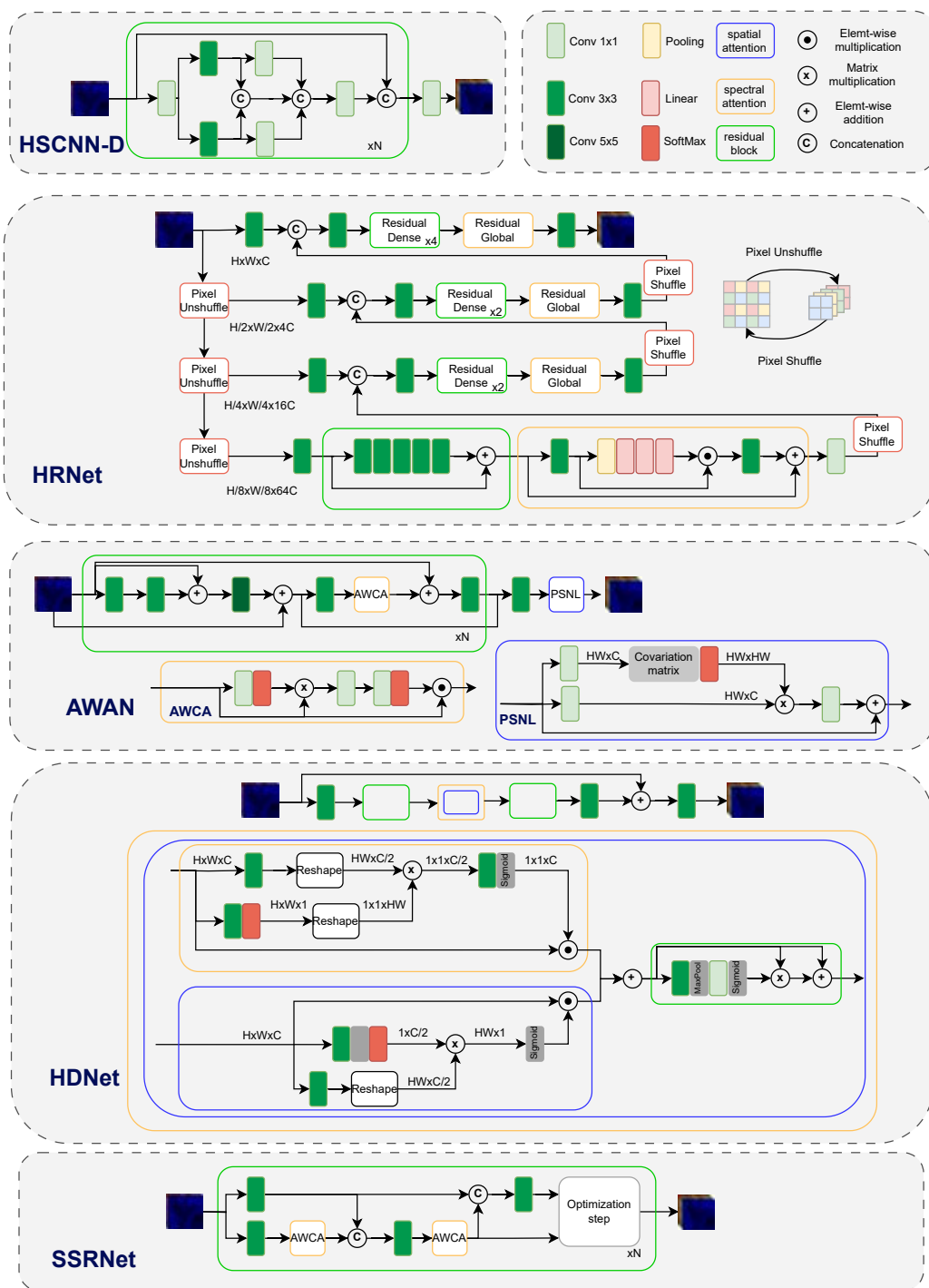


Рис. 3. Схематичные иллюстрации архитектур свёрточных нейронных сетей для решения задачи ГР. Надпись $H \times W \times C$ рядом с некоторыми стрелками, обозначают размер изображения или тензора в текущем месте нейронной сети. По умолчанию, все тензоры имеют размер $H \times W \times d$, где H , W — пространственные размеры входного изображения, d — количество спектральных каналов или размерность внутреннего представления.

внутренних признаков сети в итоговое ГСИ используются свёрточные слои с ядром размера 3×3 . В качестве функции потерь авторы использовали MRAE.

CESST. Авторы CESST [35] продолжают идеи MST++, MSFN и SPECAT по использованию U-Net с трансформерами для решения задачи ГР. Они предлагают выделять начальные признаки из каждого канала входного изображения с помощью U-Net с механизмом спектрального и пространственного внимания. Далее эти признаки подаются в слой перекрестного внимания между каналами. В качестве функции потерь использовалась сумма MRAE, SSIM и L_1 между реконструированным и реальным ГСИ для разных пространственных разрешений изображений.

PhySpec. Ян и др. [37] предлагают обучать многомодульную нейронную сеть в несколько этапов. Нейронная сеть состоит двух модулей: один производит динамическую оценку чувствительностей и спектра освещения в сцене, второй делает оценку самого ГСИ. Модуль, оценивающий ГСИ, параметризуется архитектурой U-Net, представленной в MST++ [19]. В отличие от остальных работ, авторы обучают нейронную сеть восстанавливать черные метамеры (от англ. black metamers) [55, 56, 57] — компоненты спектров, проецирующиеся в ноль при данной чувствительности камеры C . Обучение такой нейронной сети производится в 3 этапа:

1. обучение всех модулей сразу на восстановление входного RGB-изображения, используя оценку чувствительности и восстановленного ГСИ, и на восстановление ГСИ. В обоих случаях используется L_1 функция потерь;
2. дообучение вспомогательного модуля оценки чувствительности и освещения;
3. дообучение U-Net во время этапа тестирования. На данном этапе сеть дообучается так, чтобы RGB-проекция восстановленного ГСИ была максимально близка к входному RGB.

Диффузионные нейронные сети.

LDRNet. Ли и др. [36] предлагают использовать диффузионные нейронные сети для решения целого спектра задач улучшения ГСИ: шумоподавление, увеличение пространственного и спектрального разрешения. В основе предложенного решения лежит U-Net, на каждом уровне которого обработку признаков выполняют две параллельные ветки: диффузионная для уточнения спектральной информации и трансформерная для извлечения пространственных зависимостей. Трансформерная ветвь состоит из двух параллельных блоков обработки пространственного внимания с прямоугольными окнами размера $\frac{hw}{w'h'}w' \times h' \times C$, $\frac{hw}{h'w'}h' \times w' \times C$, здесь h', w' — пространственные размеры окна. Диффузионная ветвь используется для выделения из изображения малорангового словаря репрезентативных спектров, который потом используется для пересчета спектрального внимания. LDRNet обучается в три стадии:

1. обучение всей нейронной сети на задачу улучшения ГСИ с использованием реального ГСИ для выделения малорангового словаря спектров, функция потерь — L_1 ;
2. обучение диффузионного модуля для выделения словаря без контекста с реальным ГСИ с функцией потерь L_2 между словарём, выделяемым на текущем этапе, и словарём, выделенным на первом этапе;
3. дообучение U-Net с уже обученным диффузионным модулем.

4. НАБОРЫ ГИПЕРСПЕКТРАЛЬНЫХ ДАННЫХ

Хороший набор данных для решения задачи ГР должен содержать гиперспектральные фотографии объектов с разнообразными спектральными характеристиками. Эта вариативность достигается съёмкой объектов как естественного, так и искусственного происхождения, выбором разнообразных условий освещения.

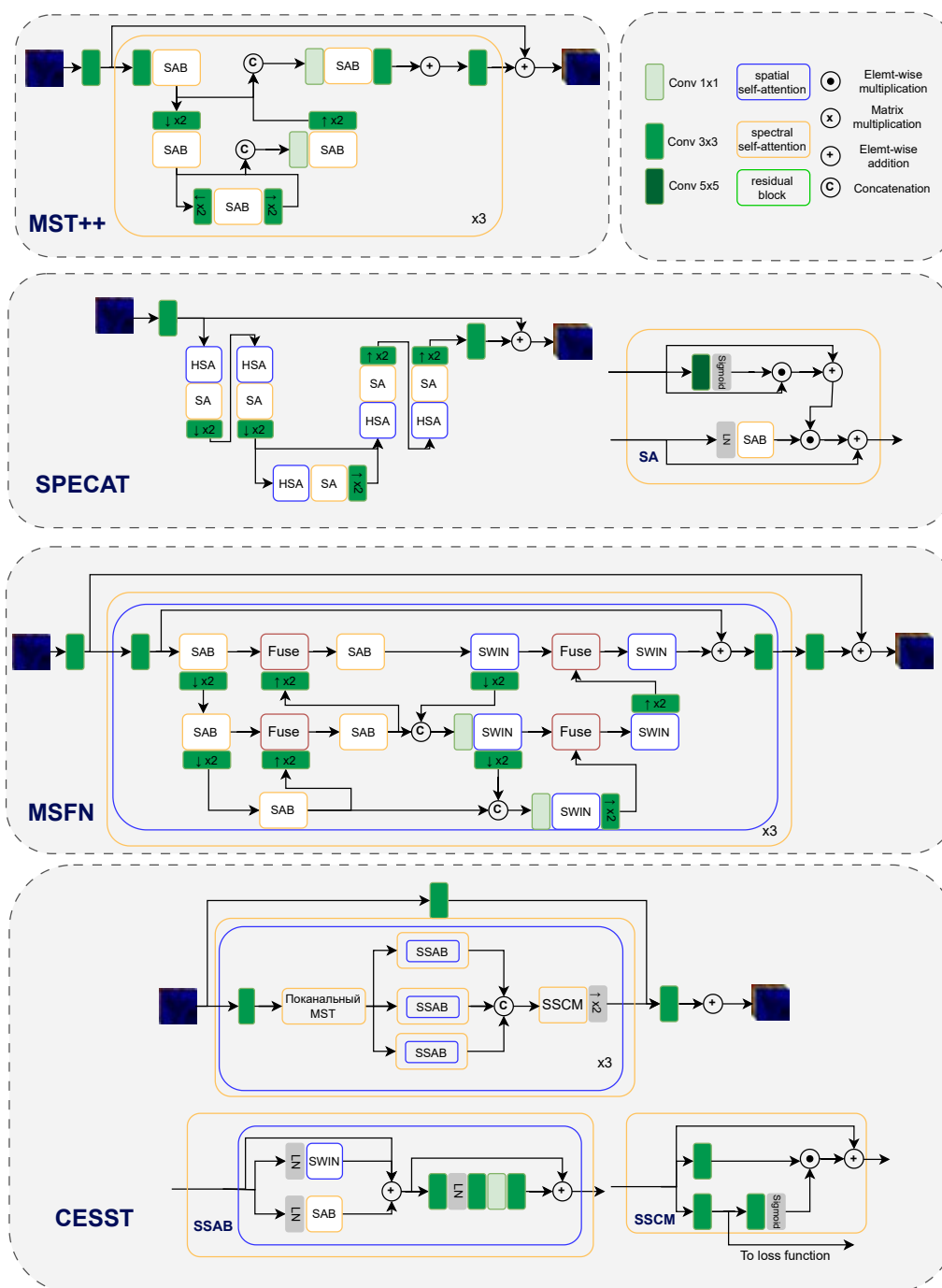


Рис. 4. Схематичные иллюстрации архитектур нейронных сетей для решения задачи ГР, использующих трансформеры. SAB — блок спектрального самовнимания (от англ. Spectral Self-Attention Block), SA — спектральное самовнимание (от англ. Spectral Self-Attention), HSA — иерархическое пространственное самовнимание (от англ. Hierarchical Spatial Self-Attention), SSAB — блок спектрально-пространственного самовнимания (от англ. Spectral-Spatial Self-Attention Block), LN — Layer Normalization.

Одним из первых таких наборов данных был CAVE [58], содержащий 32 фотографии, полученных в контролируемых лабораторных условиях. В каждом кадре присутствует белая мишень, обладающая равномерной отражательной способностью. При пост-обработке спектр каждого пикселя был поделен на спектр белой мишени, таким образом, ГСИ в наборе CAVE содержат оценки отражательных способности объектов — окраски. Такие данные подходят для симуляции цветовых откликов объектов при разных условиях освещения.

Два крупных современных набора данных — ICVL [22] и ARAD 1K [20], иногда по-другому именуемые BGU и NTIRE 2022 соответственно, содержат 200 и 950 изображений соответственно. Сбор наборов данных такого размера стал возможен благодаря появлению доступных портативных гиперспектральных камер от компании Specim. ГСИ в этих наборах преимущественно содержат фотографии сцен при дневном освещении и содержат спектры-стимулы объектов. В предложенных сценах нет эталонного объекта для оценки освещения, что ограничивает возможность оценки спектров-окрасок. Оба набора использовались в международных соревнованиях по спектральной реконструкции из RGB-изображений [38, 39, 20] в рамках конференции CVPR, за счет чего обрели большую популярность в научном сообществе.

Другой набор ГСИ — KAUST [59] — по объёму сопоставим с ICVL и ARAD 1K, содержит 409 изображений, но отличается по характеру содержания. Изображения в нем содержат окраски объектов, полученных по той же схеме, что и в CAVE. Данный набор данных обычно используется для исследования изменения цветовых откликов при изменениях условий освещения — задача цветовой константности, а так же для исследования окрасок объектов, встречающихся в повседневной жизни.

Важной характеристикой наборов данных является наличие RGB-версий соответствующих ГСИ. Дело в том, что синтезировать RGB-изображение из ГСИ можно многими способами: неоднозначность начинается уже с момента выбора матрицы чувствительности C и далее преследует исследователя при выборе параметров шума, конечного формата изображения и т. д. Подробно это было обсуждено в конце Раздела 2. Явное задание набора входных изображений позволяет надёжно стандартизировать входные данные для методов ГР и окончательно обеспечивает честное сравнение методов между собой.

В CAVE авторы предоставляют RGB-изображения спроецированные в цветовое пространство sRGB под стандартным освещением D65 в формате BMP. В ARAD 1K и ICVL предоставленные ГСИ спроецированы в цветовое пространство камеры Basler ace 2 и цветовое пространство стандартного наблюдателя CIE XYZ соответственно. RGB-изображения сохранены в формат JPG без предварительной конвертации в цветовое пространство sRGB. В KAUST авторы не предоставляют RGB-изображений.

Сводная информация о рассмотренных наборах данных собрана в Таб. 4.

Набор данных	Год	# сцен	Пространственное разрешение	Тип спектральных данных	Спектральный диапазон (Шаг дискретизации), нм	Соответствующие RGB
CAVE [58]	2010	32	512×512	Окраски	400–700 (10)	Симулированные BMP
ICVL [22]	2016	200	1392×1300	Спектры-стимулы	400–1000 (10)	Симулированные JPEG
KAUST [59]	2021	409	512×512	Окраски	400–730 (10)	Отсутствуют
ARAD 1K [20]	2022	950	512×482	Спектры-стимулы	400–700 (10)	Симулированные JPEG

Таблица 1. Сравнительная табличка, существующих бенчмарков для ГР

5. СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ

5.1. Формулировка вычислительных экспериментов

В данной работе предлагается рассмотреть 2 постановки задачи ГР:

1. реконструкция спектров-стимулов из RGB-изображений, сохранённых в формате JPG. Именно такие RGB-изображения предоставлены в наборе данных ARAD 1K. Полное описание процедуры генерации дано в [20]. Отметим, что хотя авторы старались приблизить процедуру к настоящему конвейеру обработки изображений, в ней отсутствует важный этап перехода из цветового пространства камеры в sRGB;
2. реконструкция из RAW RGB, синтезированных со спектральными чувствительностями камеры Nikon D700. Подобный подход использовался в работах [32, 36]. Дополнительно будет учтён шум, возникающий при съёмке в реальных условиях, по схеме из [20], и дискретный характер сигналов, регистрируемых камерой — представление интенсивности в виде целого числа от 0 до 2^{12} ;

Для проведения экспериментов использовались 2 набора данных: ARAD 1K и CAVE. Для набора данных ARAD 1K есть общепринятое разбиение на обучающую и валидационную выборки в соотношении 900:50. Обучение всех методов реконструкции проводилось на обучающей части набора данных ARAD 1K, а для тестирования использовались валидационная часть ARAD 1K и набор данных CAVE. Так как ГСИ в CAVE содержат отражательные способности объектов, каждое ГСИ было заменено на 4 «освещённые» версии. Освещение осуществлялось умножением окрасочного спектра в каждом пикселе на спектр одного из 4 стандартных источников: A, D65, L-15, L-RGB1. Если для реконструкции из JPG входные данные набора данных ARAD 1K использовались в предоставленном авторами формате, то для CAVE они генерировались отдельно по схеме, описанной в [20].

Для оценки качества реконструкции использовались четыре метрики:

$$\text{PSNR}(\mathbf{Y}, \hat{\mathbf{Y}}) = 10 \log_{10} \frac{(\max \mathbf{Y})^2}{\frac{1}{hwn} \sum_{i,j=1}^{h,w} \|\mathbf{y}_{ij} - \hat{\mathbf{y}}_{ij}\|_2^2}, \quad (12)$$

$$\text{SAM}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{180^\circ}{\pi} \cdot \arccos \frac{\langle \mathbf{y}, \hat{\mathbf{y}} \rangle}{\|\mathbf{y}\|_2 \cdot \|\hat{\mathbf{y}}\|_2}, \quad (13)$$

$$\text{MRAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}, \quad (14)$$

$$\text{NSE}(\mathbf{y}, \hat{\mathbf{y}}) = 100\% \cdot \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_1}{\|\mathbf{y}\|_1}, \quad (15)$$

где подсчёт метрик SAM, MRAE и NSE для пар изображений подразумевает взятие среднего по всем пикселям: $f(\mathbf{Y}, \hat{\mathbf{Y}}) := \frac{1}{hw} \sum_{i,j=1}^{h,w} f(\mathbf{y}_{ij}, \hat{\mathbf{y}}_{ij})$, $f \neq \text{PSNR}$.

PSNR — общепринятая метрика, показывающая качество восстановления интенсивности сигнала [34, 33, 36, 37]. Spectral Angle Mapper (SAM) — угол в градусах между векторами, позволяет сравнивать *формы* спектров независимо от их интенсивности [27, 32, 34, 35, 36, 37]. Mean Relative Absolute Error (MRAE) — величина, измеряющая *относительную* ошибку в каждом пикселе в каждом спектральном канале. Легко интерпретируемая метрика, однако чрезмерно штрафующая реконструкцию на тёмных, и потому шумных, участках изображения. Наряду с RMSE используется в качестве основной метрики в большинстве работ, посвященных ГР [19, 30, 28, 24, 25, 34, 32] и на международных соревнованиях [38, 39, 20]. Normalized Spectral Error (NSE) — метрика схожая с MRAE, но носит интегральный характер. Может быть проинтерпретирована как отношение площади перекрытия графиков спектров к площади под графиком реального спектра [60].

Среди других метрик, встречающихся в работах по решению задачи ГР, можно отметить RMSE, MAE [22, 24, 21, 38, 39, 20], SSIM [61, 34, 33, 32, 36, 37], его предшественника UIQI [62, 32] и EGRAS [63, 35]. RMSE и MAE измеряют ошибку в единицах измерения интенсивности спектров, а потому плохо интерпретируемы, хотя применимы для сравнения методов ГР в пределах

одного набора данных. Иными словами, выбранные нами метрики удовлетворяют свойству $f(\alpha\hat{\mathbf{Y}}, \alpha\mathbf{Y}) = f(\hat{\mathbf{Y}}, \mathbf{Y})$, тогда как RMSE и MAE нет. SSIM, UIQI и EGRAS — метрики структурной схожести изображений, более подходящие для задач создания визуально-приемлемого выхода: шумоподавление, повышение пространственного разрешения. Также помимо схожести восстановленного и реального ГСИ иногда используются метрики схожести RGB-проекций этих изображений [57, 37].

5.2. Качественные и количественные результаты гиперспектральной реконструкции

Для обучения методов реконструкции гиперпараметры, оптимизаторы и стратегии изменения скорости обучения (от англ. learning rate) выбирались такими же, как в оригинальных статьях.

Результаты реконструкции из JPG RGB и из RAW RGB представлены в Таб. 2 и 3 соответственно. Из таблиц видно, что реконструкция из RAW RGB позволяет улучшить качество оценки спектров в среднем на 20% по сравнению с реконструкцией из JPG RGB вне зависимости от используемого метода. Улучшение метрик PSNR и NSE значительно лучше, в среднем на 20 %, чем метрики SAM, в среднем 10 %. Стоит заметить, что аналитические методы реконструкции спектров показывают качество сравнимое с нейронной сетью HSCNN+ и работают стабильно хуже более поздних методов на основе нейронных сетей, если тестировались на изображениях из того же набора данных, на котором учились. При тестировании же на другом наборе данных они способны обгонять методы на основе нейронных сетей, что свидетельствует о лучшей обобщающей способности первых.

Большие значения NSE на наборе данных CAVE объясняются высокой представленностью тёмных регионов в изображениях оттуда.

Метод реконструкции	NTIRE 2022 (тренировочный датасет)			Освещенный CAVE (датасет для валидации)		
	PSNR, dB ↑	SAM, ° ↓	NSE, % ↓	PSNR, dB ↑	SAM, ° ↓	NSE, % ↓
Linreg	19,17	8,30	64,87	23,19	28,61	285,17
A+	20,00	8,84	45,9	21,65	31,65	233,26
RBFN	19,25	7,42	70,37	21,55	31,28	301,96
HSCNN+	24,72	8,86	51,15	19,61	28,37	610,37
HRNet	25,08	5,80	39,84	19,56	29,72	667,73
AWAN	30,41	6,05	21,28	20,47	31,40	609,72
HDNet	28,15	6,89	27,34	22,20	31,03	465,98
MST++	32,00	6,05	16,66	19,24	28,41	705,60
SSRNet	27,61	7,34	29,00	21,68	29,27	488,82
SPECAT	30,71	5,33	22,22	18,74	28,30	764,89
MSFN	29,93	4,76	21,91	21,57	28,79	535,90
CESST	31,69	5,73	17,19	17,76	29,19	897,53

Таблица 2. Результаты гиперспектральной реконструкции из JPG RGB. Жирным выделено лучшее значение метрики в столбце, подчеркнутым — второй лучший результат.

Для качественного анализа реконструкции на Рис. 5, 6, 8, 9, приводятся карты ошибок SAM и NSE. Обратим внимание на следующие аспекты:

1. на картах SAM видны артефакты сжатия JPG, например, у метода HSCNN+ Рис. 6;
2. на картах ошибок в случае реконструкции из RAW RGB, Рис. 8 и 9, видны зернистые шумовые артефакты у попиксельных методов и HSCNN+, эти методы объединяет отсутствие анализа пирамидальной структуры изображения, так, у более современных методов реконструкции на основе нейронной сети карты ошибок выглядят более сглаженными;

Метод реконструкции	NTIRE 2022 (тренировочный датасет)			CAVE (датасет для валидации)		
	PSNR, dB ↑	SAM, ° ↓	NSE, % ↓	PSNR, dB ↑	SAM, ° ↓	NSE, % ↓
Linreg	26,55	9,85	35,21	26,38	31,16	53,14
A+	26,32	9,61	28,70	24,97	34,68	59,32
RBFN	26,69	8,63	33,16	25,90	30,73	62,93
HSCNN+	27,0	6,50	32,88	24,02	27,89	114,68
HRNet	27,85	4,66	26,30	24,12	28,17	113,82
AWAN	31,90	5,32	16,28	25,96	28,47	126,33
HDNet	30,71	6,43	19,19	24,44	31,78	92,56
MST++	31,49	5,24	17,48	27,50	21,08	74,80
SSRNet	30,49	5,76	18,99	25,17	29,75	120,87
SPECAT	33,52	4,34	13,61	25,22	27,60	124,94
MSFN	29,49	4,60	21,49	25,98	22,83	86,70
CESST	31,76	5,03	17,54	25,50	27,93	116,11

Таблица 3. Результаты гиперспектральной реконструкции из RAW RGB. Жирным выделено лучшее значение метрики в столбце, подчеркнутым — второй лучший результат.

- стоит заметить, что у всех методов реконструкции ошибки в тёмных участках сцены выше чем на светлых;
- также ошибки высокие на границах объектов.

Примеры восстановленных спектров для реконструкции из JPG RGB и из RAW RGB приведены на Рис. 7 и 10 соответственно.

6. ОЦЕНКА НАТУРАЛЬНОСТИ ВОССТАНАВЛИВАЕМЫХ СПЕКТРОВ

Важным аспектом при работе с реконструированными ГСИ является физическая реалистичность и корректность полученных данных. Под этим подразумевается возможность встретить восстановленный спектр в реальной жизни. В описанных до этого экспериментах не накладывались условия на восстанавливаемый спектр, из-за этого методы реконструкции могут сгенерировать несуществующий или далекий от натуральных спектр. Данная проблема рассматривалась в работе Фсиана и др. [64]. Авторы изучали распределения восстановленных и реальных гиперспектров, спроецировав их на первые две главные компоненты PCA. Их эксперименты показывают, что в части случаев распределения не совпадают, причем распределение восстановленных спектров не вкладывается в распределение реальных, а иногда даже не пересекается с ним, в том числе для передового метода ГР того времени — MST++.

Данный эксперимент предлагается улучшить, дав количественную оценку расстояния между распределениями спектров. Для этого предлагается посчитать расстояние Синхорна [65] в пространстве спектров между распределением реальных и восстановленных. Значения метрики расстояния Синхорна для реконструкции из JPG RGB представлены в Таб. 5, для реконструкции из RAW RGB в Таб. 4. Стоит заметить, что расстояния между распределениями у классических методов лучше при восстановлении из RAW RGB, примерно на 11%. У методов на основе нейронных сетей в большинстве случаев лучше при восстановлении из JPG примерно на 2%.

Для качественной оценки совпадения распределений спектров они были спроецированы на две главные компоненты, выделенные с помощью PCA. Для такого анализа были выбраны по одному методу реконструкции из разных классов: A+ — словарный попиксельный метод, AWAN — СНС с механизмом спектрального внимания, MST++ — трансформер со спектральным самовниманием и CESST — трансформер с механизмом спектрального и пространственного самовнимания. Проекция распределений спектров представлены на рисунке 11. Для по-

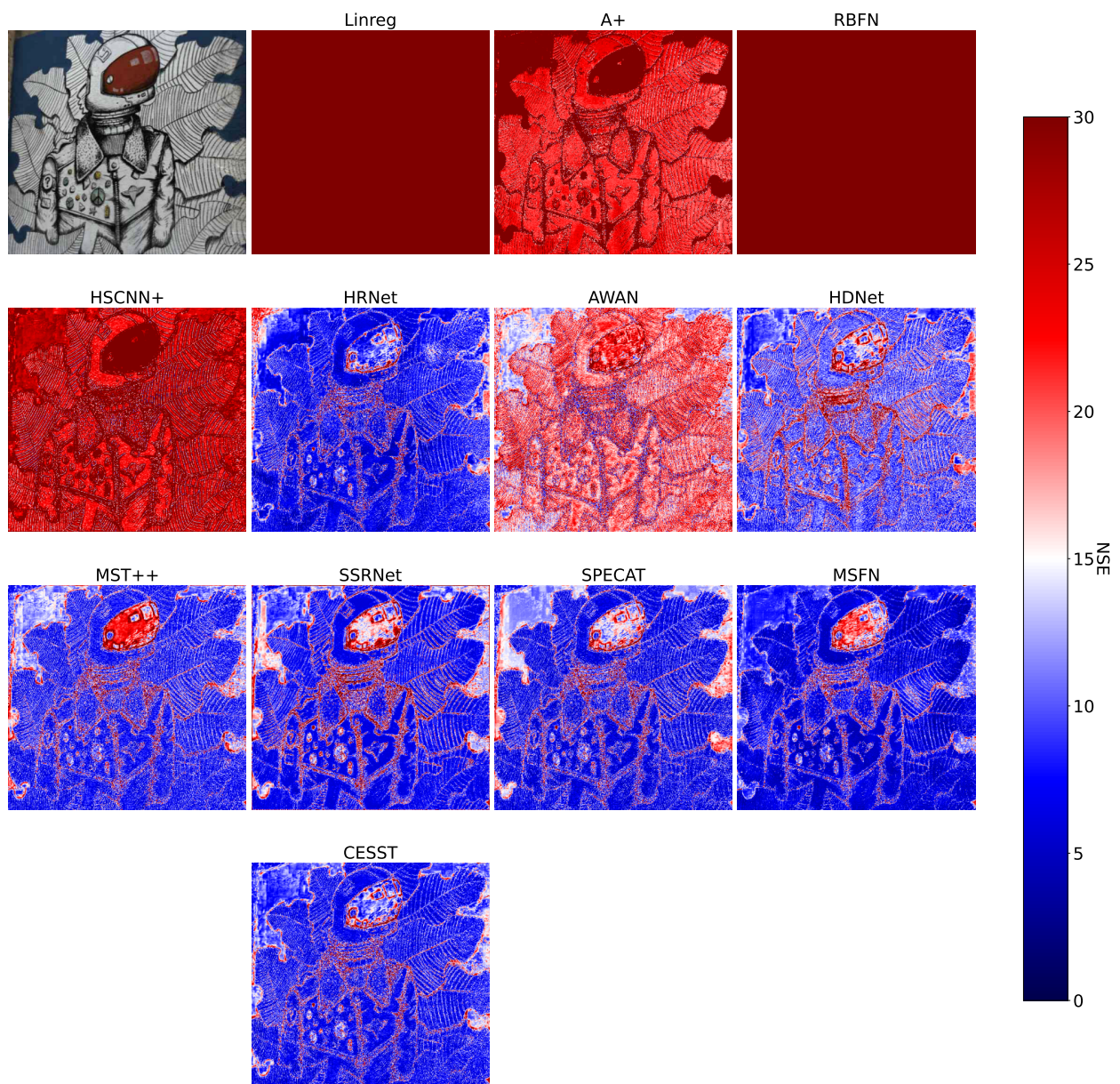


Рис. 5. Карта ошибок NSE методов реконструкции, обученных на JPG RGB. Изображение взято из валидационной части набора данных ARAD 1K.

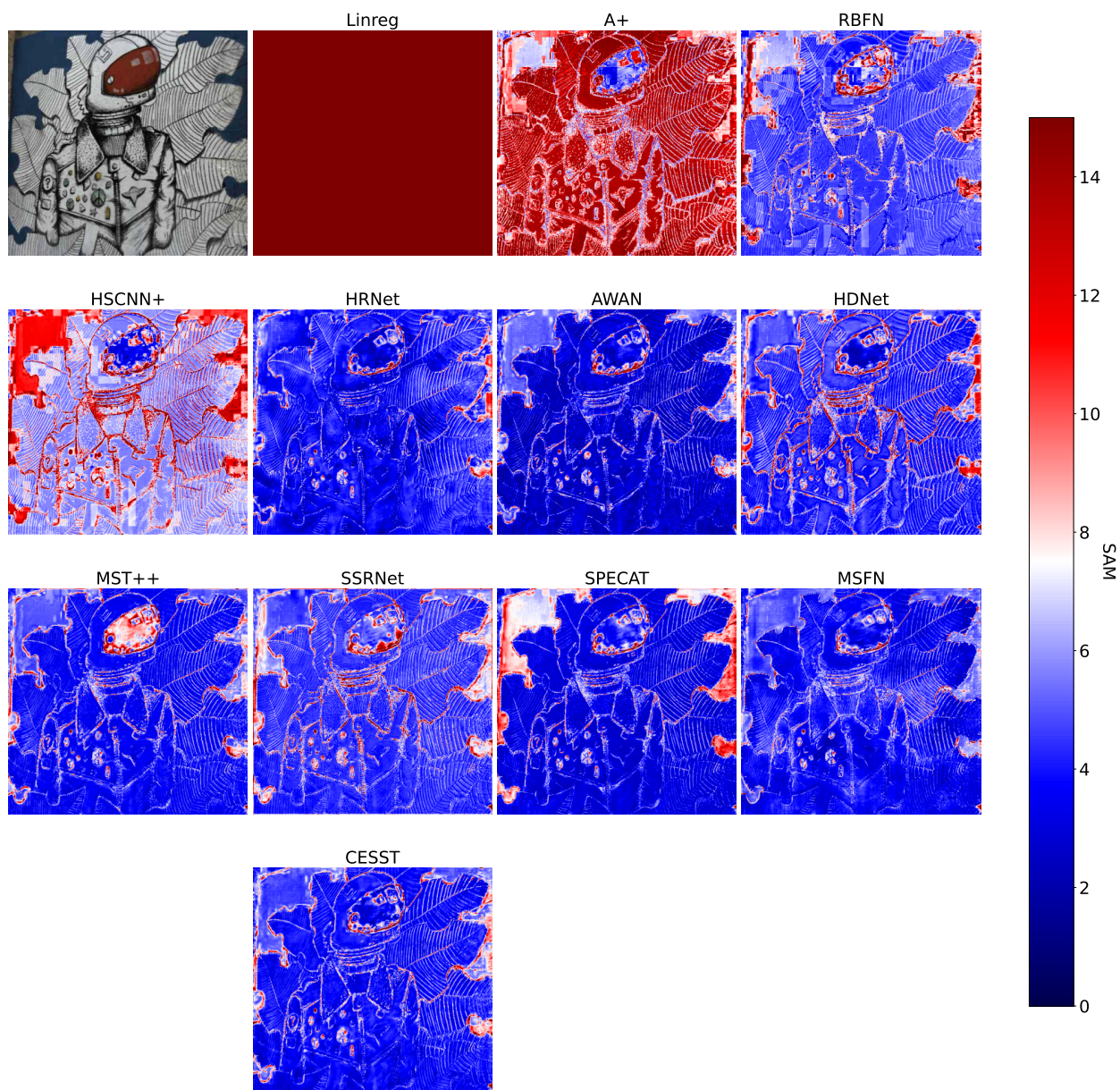


Рис. 6. Карта ошибок SAM методов реконструкции, обученных на JPG RGB. Изображение взято из валидационной части набора данных ARAD 1K.

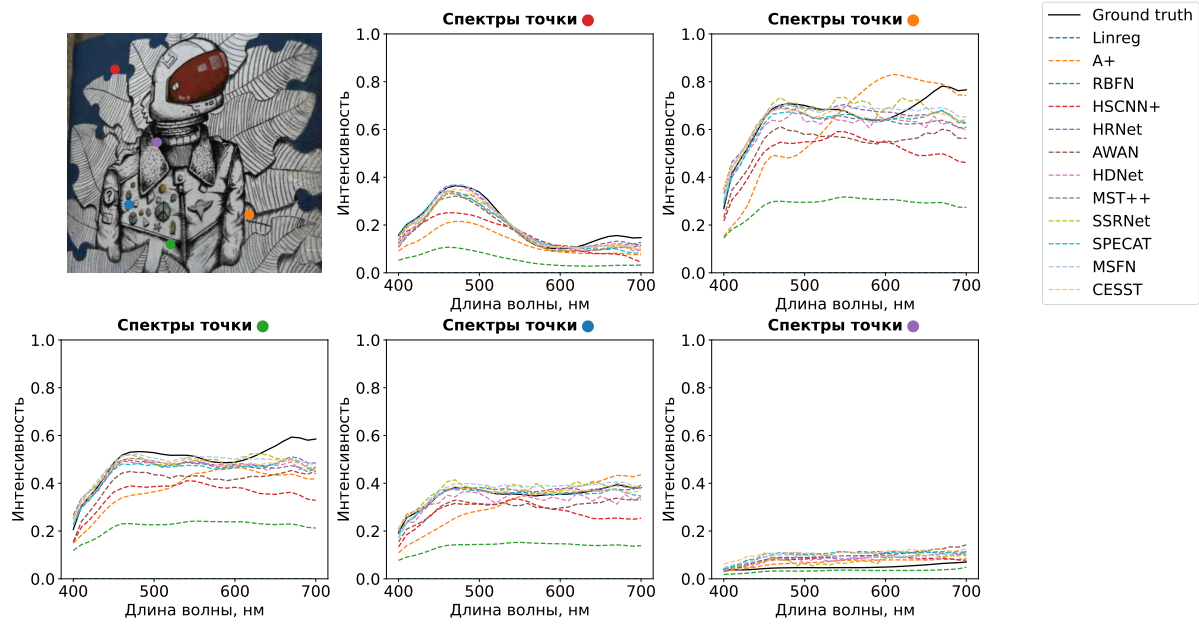


Рис. 7. Примеры реконструкции спектров методами реконструкции, обученных на JPG RGB. Изображение взято из валидационной части набора данных ARAD 1K.

Метод реконструкции	NTIRE (тренировочный датасет)	CAVE (датасет для валидации)
	Расстояние Синхорна ↓	Расстояние Синхорна ↓
Linreg	0.0873	0.4909
A+	0.1051	0.5200
RBFN	0.0879	0.5511
HSCNN+	0.0809	0.2689
HRNet	0.0547	0.2534
AWAN	0.0647	0.2392
HDNet	0.0722	0.2119
MST++	0.0618	0.2645
SSRNet	0.0707	<u>0.2309</u>
SPECAT	0.0522	0.2691
MSFN	<u>0.0523</u>	0.2388
CESST	0.0615	0.2945

Таблица 4. Расстояние Синхорна между распределениями восстановленных и реальных спектров на валидационной части набора данных ARAD 1K и набора данных CAVE для методов, обученных восстанавливать спектры-стимулы из JPG RGB. Жирным выделено лучшее значение метрики в столбце, подчёркнутым — второй лучший результат.

строения графиков использовались методы реконструкции, обученные на JPG RGB. У метода A+ есть выраженное распределение вдоль первой главной компоненты и спектры, которые он аппроксимирует, не совпадают с большим облаком реальных спектров. Нейросетевые методы же лучше аппроксимируют маловероятные реализации спектра, они лучше приближают отдельные малые облака точек и основное большое облако точек.

7. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ И НЕРЕШЕННЫХ ЗАДАЧ В СФЕРЕ ГР

Современные методы реконструкции на основе нейронных сетей с механизмами самовнимания показывают лучшие результаты реконструкции в сравнении со свёрточными нейронными

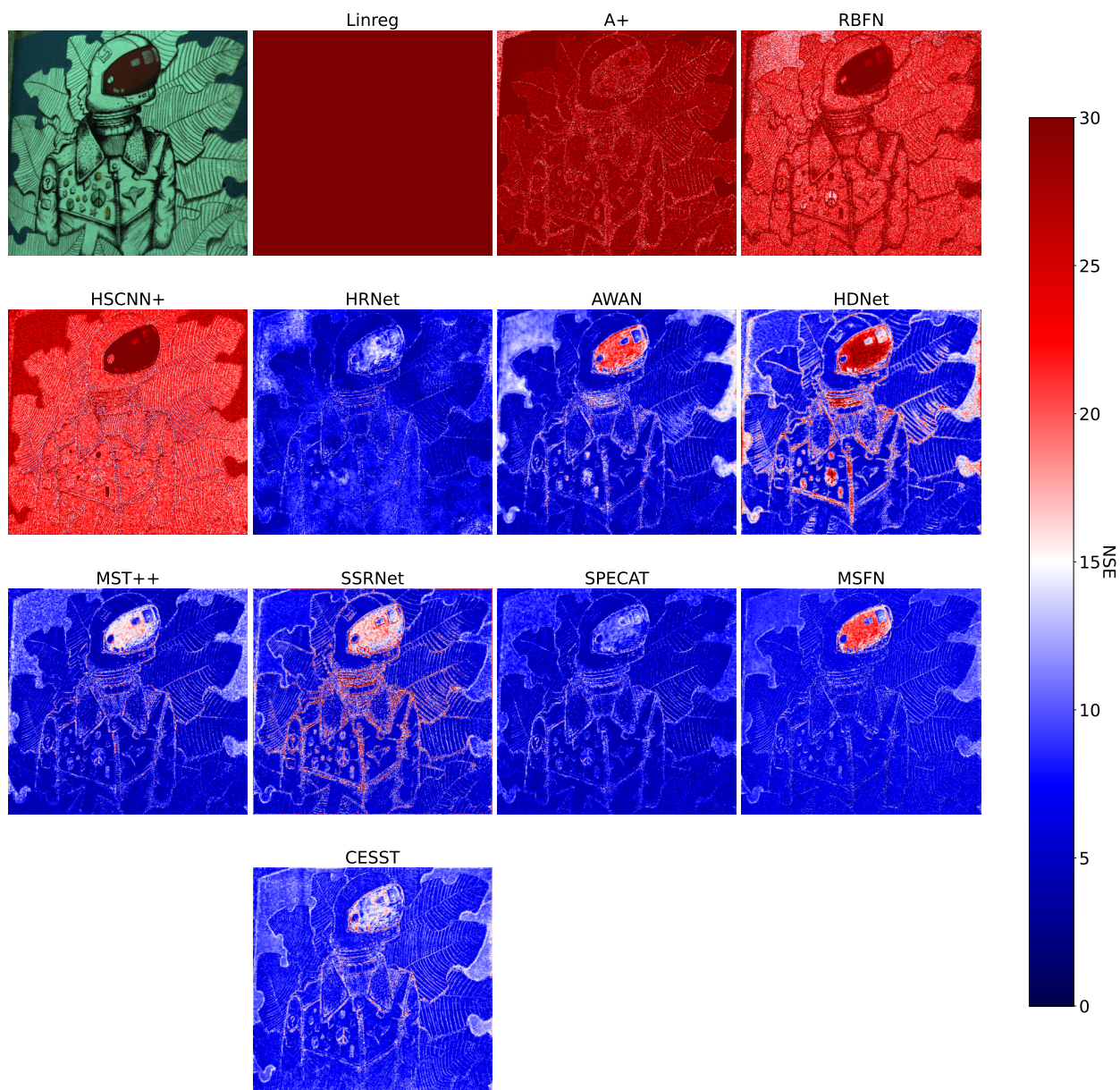


Рис. 8. Карта ошибок NSE методов реконструкции, обученных на RAW RGB. Изображение взято из валидационной части набора данных ARAD 1K.

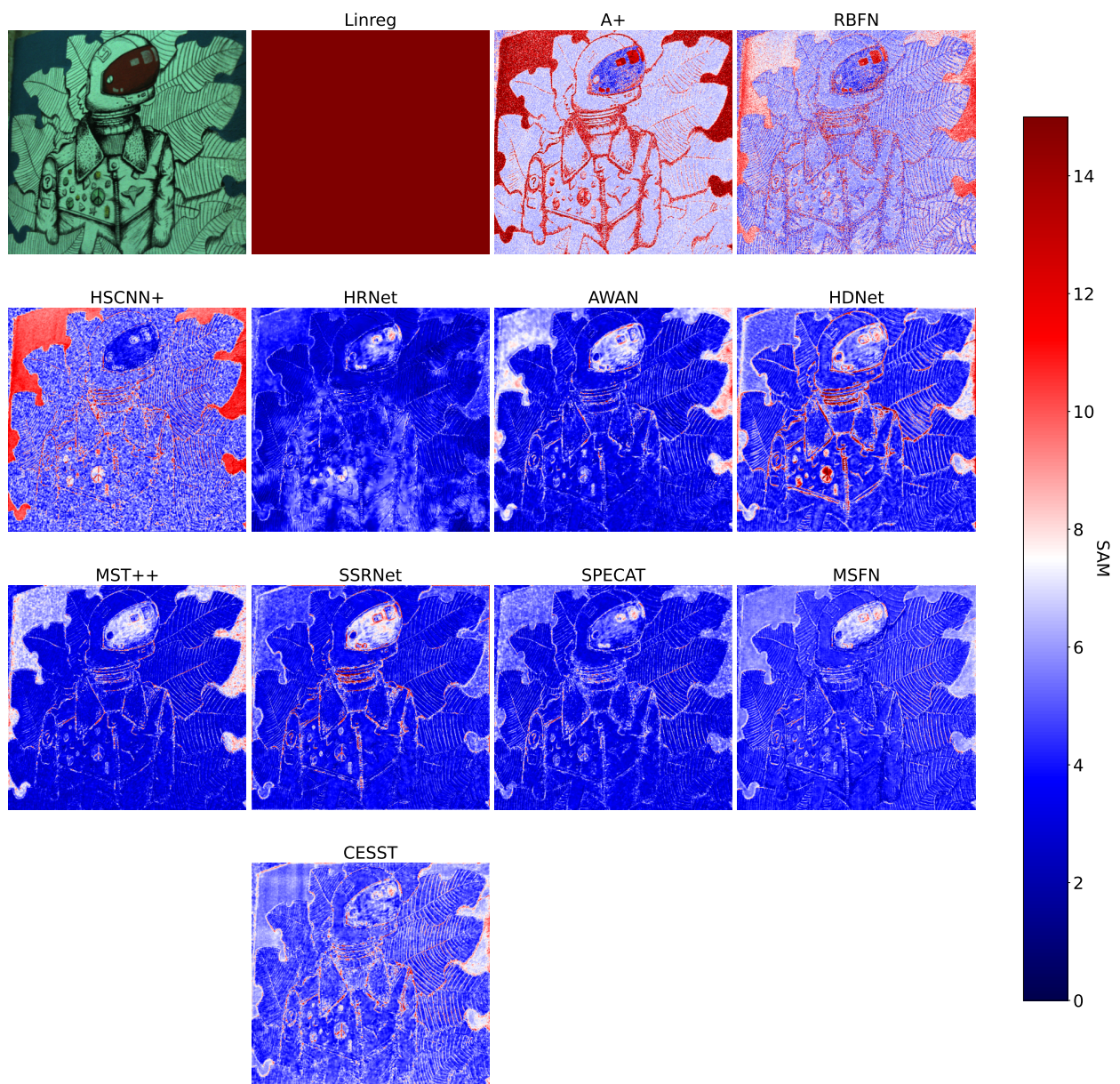


Рис. 9. Карта ошибок SAM методов реконструкции, обученных на RAW RGB. Изображение взято из валидационной части набора данных ARAD 1K.

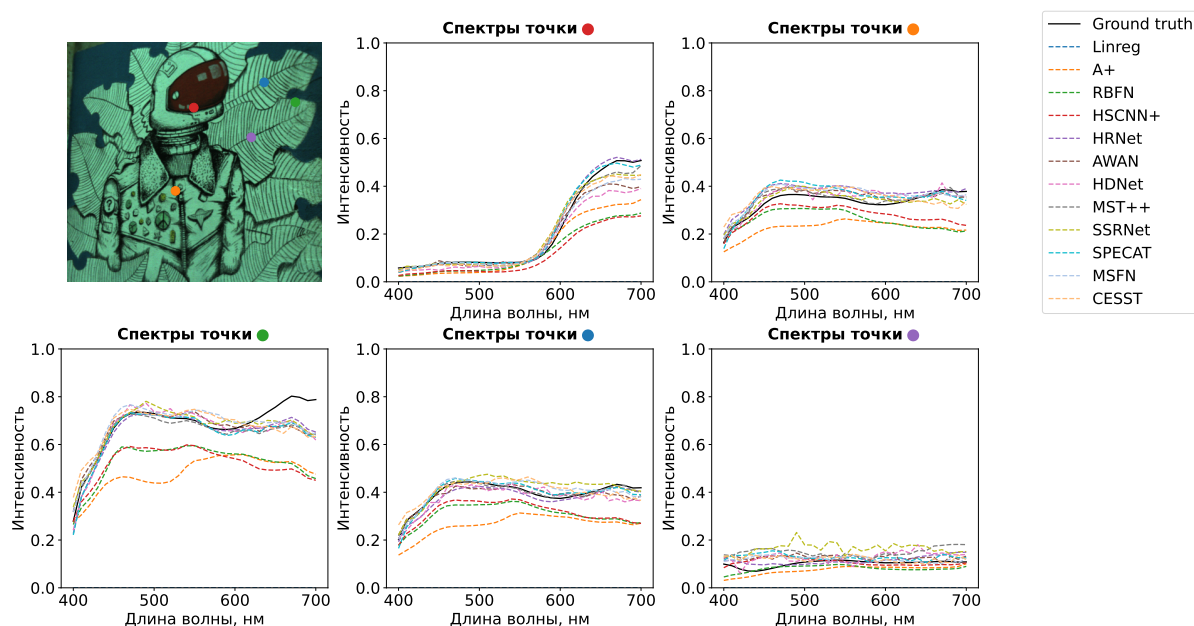


Рис. 10. Примеры реконструкции спектров методами реконструкции, обученных на RAW RGB. Изображение взято из валидационной части набора данных ARAD 1K.

Метод реконструкции	NTIRE (тренировочный датасет)	CAVE (датасет для валидации)
	Расстояние Синхорна ↓	Расстояние Синхорна ↓
Linreg	0.0926	0.4909
A+	0.1108	0.5200
RBFN	0.1367	0.5511
HSCNN+	0.0948	0.1762
HRNet	0.0604	0.1742
AWAN	<u>0.0591</u>	0.1668
HDNet	0.0666	0.1738
MST++	0.0607	<u>0.1632</u>
SSRNet	0.0666	0.1739
SPECAT	0.0534	0.1669
MSFN	0.0688	0.1439
CESST	0.0612	0.1686

Таблица 5. Расстояние Синхорна между распределениями восстановленных и реальных спектров на валидационной части набора данных ARAD 1K и набора данных CAVE для методов, обученных восстанавливать спектры-стимулы из RAW RGB. Жирным выделено лучшее значение метрики в столбце, подчёркнутым — второй лучший результат.

сетями и не нейросетевыми алгоритмами. При этом трансформерные методы реконструкции используют сравнительно небольшое количество параметров, порядка миллионов, в то время как современные большие модели используют миллиарды. Вопрос, который возникает при обучении таких сравнительно небольших нейронных сетей на современных наборах ГСИ: сколько требуется данных и каких, чтобы их было достаточно для достижения высокой обобщающей способности? Так, самый большой открытый набор ГСИ повседневных сцен содержит 950 изображений размера 482×512 или же 234 444 800 спектральных кривых. На первый взгляд, такое количество примеров в обучающем наборе может показаться достаточно большим, однако, так как многие кривые берутся с одного изображения, они сильно скоррелированы. Данный факт также был показан в работе Фу и др. [44], в которой авторы уменьшили обучающий набор

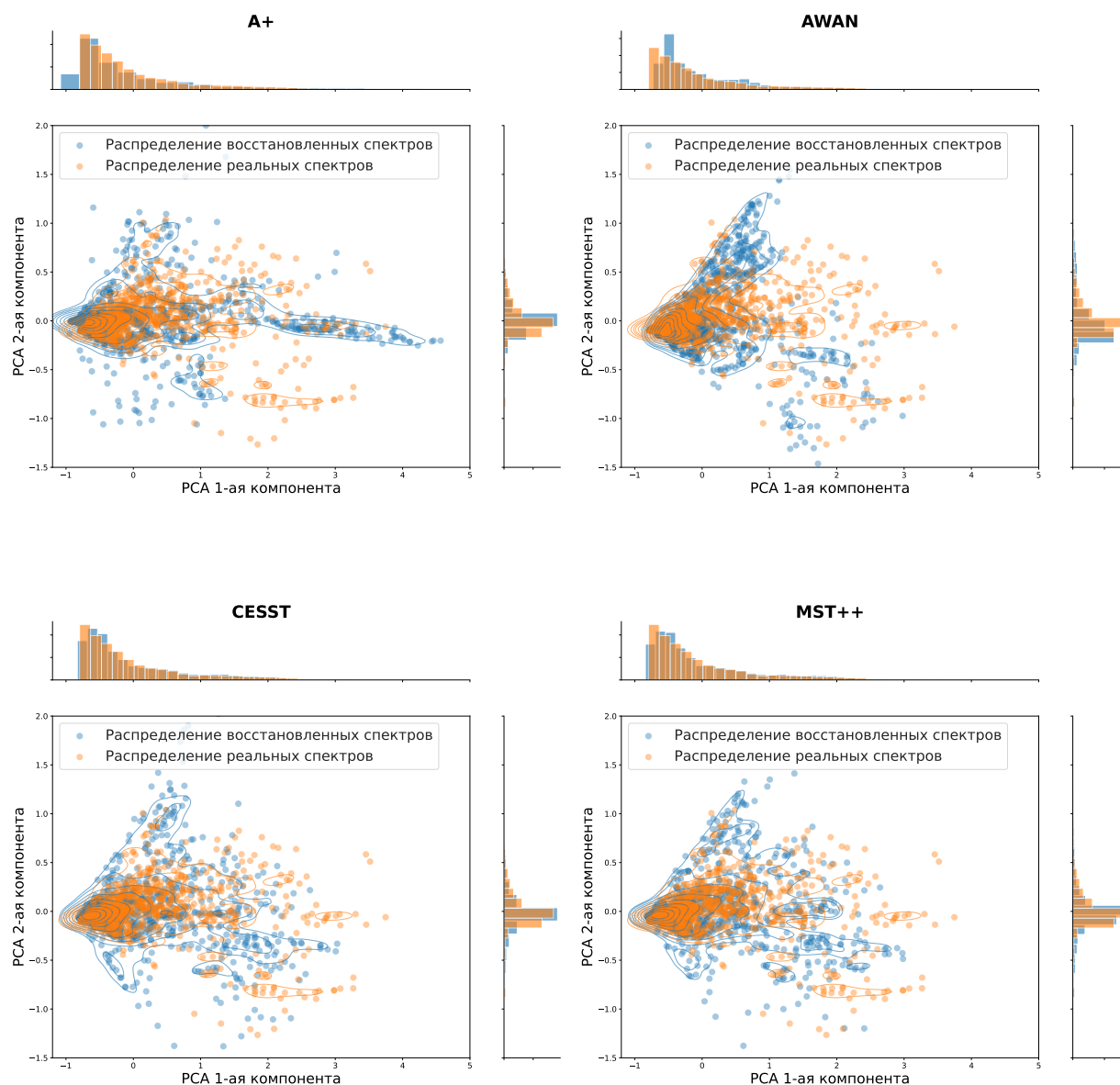


Рис. 11. Распределение проекций реальных и восстановленных спектров для методов реконструкции A+, AWAN, CESST, MST++. 2D распределения оценены с помощью KDE с гауссовским ядром.

ARAD 1K в два раза, при этом качество ГР во всех случаях ухудшилось не более чем на 20%, что подтверждает факт низкого разнообразия спектральных данных в ARAD 1K. Текущие эксперименты показывают, что такого количества данных достаточно для реконструкции спектров в сценариях съёмки, схожих с наблюдаемыми в тренировочном наборе данных, однако при других условиях качество реконструкции заметно падает. Ответ на вопрос «можно ли данные сети адаптировать под желаемый сценарий съёмки» ещё предстоит найти в будущих исследованиях.

Также в работе Фу [44] был рассмотрен вопрос устойчивости методов реконструкции к параметрам шума и силе сжатия JPG. Авторы показали, что методы ГР на основе нейронных

сетей неустойчивы к изменениям этих параметров, и переобучаются под них. При смене этих параметров качество реконструкции падает в 2 раза.

Одной из траекторий улучшения генерализации методов ГР и их устойчивости к изменениям в реальных данных может стать уточнение модели формирования изображения. В упомянутой работе Фу [44] авторы, например, предлагают симулировать хроматические аберрации при формировании изображений. Мослех и соавт. [66] предлагают непараметрическую модель шума при формировании RGB-изображения, добавляя в рассмотрение фиксированный шумовой профиль сенсора, шум квантизации и шум сжатия, что позволило наилучшим образом приблизить шум реальных сенсоров. PSNR между реальным и симулированным сигналом достигает 39dB.

Другим важным направлением развития по улучшению и генерализации методов гиперспектральной реконструкции может стать использование больших предобученных фундаментальных моделей. Одна из возможных причин, почему нейронные сети справляются со спектральной реконструкцией лучше аналитических методов, может заключаться в неявном обучении извлечению семантической информации о сцене [25]. Таким образом, для отдельных участков сцены сужается поле поиска спектральных откликов, как в работе об аппроксимации Гауссовскими процессами [27]. Как было сказано ранее, современные методы для задачи ГР содержат порядка миллиона параметров, в то время как современные нейронные сети для решения задач компьютерного зрения общего назначения, например, DINO V2 [67], DINO V3 [68], CLIP [69], содержат миллиарды параметров. Поэтому одним из направлений улучшения качества реконструкции в задаче ГР может стать использование выделенных семантических признаков из RGB-изображений, а задачу реконструкции будет решать отдельный легковесный модуль, обученный на этих признаках.

Также можно отметить тенденцию повышения спектральной размерности входных изображений. В работе [8] ГР производилась из пары RGB и инфракрасного изображений. Ох и др. [70] предлагают снимать сцену на обычные три RGB-камеры с различными спектральными чувствительностями.

8. ЗАКЛЮЧЕНИЕ

В данной работе был проведен анализ самых известных современных методов гиперспектральной реконструкции из RGB-изображений, включая последние решения на базе трансформеров и диффузионных нейронных сетей.

Был проведен анализ генерализации существующих решений, а также анализ физической правдоподобности восстанавливаемых спектров. Существующие методы реконструкции восстанавливают спектр с точностью порядка 33 dB PSNR и 4 градусов по метрике SAM, на обучающем домене. При смене тестового набора данных качество реконструкции методов на основе нейронных сетей ухудшается вплоть до 20 dB PSNR и 30 градусов SAM. При этом не нейросетевые методы реконструкции показывают лучшую обобщающую способность. Также нужно заметить, что существующие решения могут генерировать неестественные спектры, которых не было в обучающей выборке.

Все протестированные решения были объединены в единую открытую кодовую базу. Были описаны возможные альтернативные направления исследований ГР — использование более качественной спектральной аппаратуры, использование больших предобученных нейронных сетей для улучшения понимания семантики сцены или разработка более физически точной модели формирования цветового изображения. Главным вызовом на данный момент является разработка метода реконструкции устойчивого как к смене домена входных данных, так и к смене домена ГС данных.

СПИСОК ЛИТЕРАТУРЫ

1. Meiling Wang, Yongchang Xu, Zhisheng Wang, and Changda Xing. Deep margin cosine autoencoder-based medical hyperspectral image classification for tumor diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 72:1–12, 2023.
2. Xiang Xu, Jun Li, Xin Huang, Mauro Dalla Mura, and Antonio Plaza. Multiple morphological component analysis based decomposition for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):3083–3102, 2016.
3. Mohammed Abdulmajeed Moharram and Divya Meena Sundaram. Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions. *Neurocomputing*, 536:90–113, 2023.
4. Seon Joo Kim, Fanbo Deng, and Michael S Brown. Visual enhancement of old documents with hyperspectral imaging. *Pattern Recognition*, 44(7):1461–1469, 2011.
5. Md Toukir Ahmed, Ocean Monjur, and Mohammed Kamruzzaman. Deep learning-based hyperspectral image reconstruction for quality assessment of agro-product. *Journal of Food Engineering*, 382:112223, 2024.
6. Da-Wen Sun, Hongbin Pu, and Jingxiao Yu. Applications of hyperspectral imaging technology in the food industry. *Nature Reviews Electrical Engineering*, 1(4):251–263, 2024.
7. Ortal Glatt. Beyond rgb: a real world dataset for multispectral imaging in mobile devices. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4344–4354, 2024.
8. Neha Sharma, Muhammad Shahzaib Waseem, Shahrzad Mirzaei, and Mohamed Hefeeda. Mobispectral: Hyperspectral imaging on mobile devices. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023.
9. Specim. Specim iq. <https://www.specim.com/iq/>. Accessed: 2025-10-24.
10. Specim. Specim fx17. <https://www.specim.com/products/specim-gx17/>. Accessed: 2025-10-24.
11. Vladislav Batshev, Alexander Machikhin, Alexey Gorevoy, Grigoriy Martynov, Demid Khokhlov, Sergey Boritko, Vitold Pozhar, and Vladimir Lomonov. Spectral imaging experiments with various optical schemes based on the same aotf. *Materials*, 14(11):2984, 2021.
12. Xieluyao Xu, Wenhao Li, Jirigalantu, Hongzhu Yu, Jiawei Zhang, and Hongzhi Liang. High-speed measurement method for diffraction efficiency of gratings in broad range wavelength based on aotf. *Optics Express*, 32(20):34887–34897, 2024.
13. Elliot S Wachman, Wen-hua Niu, and Daniel L Farkas. Aotf microscope for imaging with increased speed and spectral versatility. *Biophysical journal*, 73(3):1215–1222, 1997.
14. Chi Chen, Yongcheng Wang, Ning Zhang, Yuxi Zhang, and Zhikang Zhao. A review of hyperspectral image super-resolution based on deep learning. *Remote Sensing*, 15(11):2853, 2023.
15. Dioline Sara, Ajay Kumar Mandava, Arun Kumar, Shiny Duella, and Anitha Jude. Hyperspectral and multispectral image fusion techniques for high resolution applications: A review. *Earth Science Informatics*, 14(4):1685–1705, 2021.
16. Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express*, 15(21):14013–14027, 2007.
17. Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51, 2008.
18. Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *European Conference on Computer Vision*, pages 187–204. Springer, 2020.

19. Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 745–755, 2022.
20. Boaz Arad. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 863–881, 2022.
21. Yan Jia, Yinqiang Zheng, Lin Gu, Art Subpa-Asa, Antony Lam, Yoichi Sato, and Imari Sato. From rgb to spectrum for natural scenes via manifold-based mapping. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4705–4713, 2017.
22. Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34, 2016.
23. David H Foster, Kinjiro Amano, Sérgio M C Nascimento, and Michael J Foster. Frequency of metamerism in natural scenes. *Journal of the Optical Society of America A*, 23(10):2359–2372, 2006.
24. Jonas Aeschbacher, Jiqing Wu, and Radu Timofte. In defense of shallow learned spectral reconstruction from rgb images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 471–479, 2017.
25. Yi-Tun Lin and Graham D. Finlayson. A rehabilitation of pixel-based spectral reconstruction from rgb images. *Sensors*, 23(8):4155, 2023.
26. ПП Николаев, СМ Карпенко, and ДП Николаев. Спектральные модели цветовой константности: правила отбора. *Труды Института системного анализа Российской академии наук*, 38:322–335, 2008.
27. Naveed Akhtar and Ajmal Mian. Hyperspectral recovery from rgb images using gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):100–113, 2018.
28. Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018.
29. Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. Hierarchical regression network for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 422–423, 2020.
30. Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, and Fei Liu. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 462–463, 2020.
31. Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17542–17551, 2022.
32. Renwei Dian, Tianci Shan, Wei He, and Haibo Liu. Spectral super-resolution via model-guided cross-fusion network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7):10059–10070, 2023.
33. Zhiyang Yao, Shuyang Liu, Xiaoyun Yuan, and Lu Fang. Specat: Spatial-spectral cumulative-attention transformer for high-resolution hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25368–25377, 2024.
34. Yaohang Wu, Renwei Dian, and Shutao Li. Multistage spatial-spectral fusion network for spectral super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 36(7):12736–12746, 2024.
35. Xingxing Yang, Jie Chen, and Zaifeng Yang. Hyperspectral image reconstruction via combinatorial embedding of cross-channel spatio-spectral clues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 6567–6575, 2024.

36. Miaoyu Li, Ying Fu, Tao Zhang, Ji Liu, Dejing Dou, Chenggang Yan, and Yulun Zhang. Latent diffusion enhanced rectangle transformer for hyperspectral image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(1):549–564, 2024.
37. Xingxing Yang, Jie Chen, and Zaifeng Yang. Physpec: Physically consistent spectral reconstruction via orthogonal subspace decomposition and self-supervised meta-auxiliary learning. In *Forty-second International Conference on Machine Learning*, 2025.
38. Boaz Arad. Ntire 2018 challenge on spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1042–1051, 2018.
39. Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham D Finlayson. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447, 2020.
40. Pai Chet Ng. Hyperspectral skin vision challenge: Can your camera see beyond your skin? In *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 59–60. IEEE, 2024.
41. Jingang Zhang, Runmu Su, Qiang Fu, Wenqi Ren, Felix Heide, and Yunfeng Nie. A survey on computational spectral reconstruction methods from rgb to hyperspectral imaging. *Scientific Reports*, 12(1):11905, 2022.
42. Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.
43. Bernd Jahne. *Practical handbook on image processing for scientific and technical applications*. CRC press, 2004.
44. Qiang Fu, Matheus Souza, Eunsue Choi, Suhyun Shin, Seung-Hwan Baek, and Wolfgang Heidrich. Limitations of data-driven spectral reconstruction: An optics-aware analysis. *IEEE Transactions on Computational Imaging*, pages 1–16, 2025.
45. Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.
46. Naveed Akhtar, Faisal Shafait, and Ajmal Mian. Sparse spatio-spectral representation for hyperspectral image super-resolution. In *European conference on computer vision*, pages 63–78. Springer, 2014.
47. Yagyensh Chandra Pati, Ramin Rezaiifar, and Perinkulam Sambamurthy Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44, 1993.
48. Mukund Balasubramanian and Eric L Schwartz. The isomap algorithm and topological stability. *Science*, 295(5552):7–7, 2002.
49. Zhong Zhao and Guocan Feng. A dictionary-based algorithm for dimensionality reduction and data reconstruction. In *2014 22nd International Conference on Pattern Recognition*, pages 1556–1561, 2014.
50. Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, pages 689–696, 2009.
51. V.V. Maximov. Color transformation with changes of illumination, 1984.
52. Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
53. Yuanyuan Zhao, Hui Guo, Zhan Ma, Xun Cao, Tao Yue, and Xuemei Hu. Hyperspectral imaging with random printed mask. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10149–10157, 2019.

54. Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.
55. Graham D Finlayson and Peter Morovic. Metamer sets. *Journal of the Optical Society of America A*, 22(5):810–819, 2005.
56. Jozef B Cohen and William E Kappauf. Metameric color stimuli, fundamental metamers, and wyszecki's metamer black. *The American journal of psychology*, pages 537–564, 1982.
57. Yi-Tun Lin and Graham D. Finlayson. Physically plausible spectral reconstruction. *Sensors*, 20(21):6399, 2020.
58. Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253, 2010.
59. Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Multispectral illumination estimation using deep unrolling network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2672–2681, 2021.
60. D. Reutsky, A. Pogadaev, D. Vladimirov, and E. Ershov. Towards using smartphones as hyperspectral cameras. *Computer Optics*, 49(1):112–120, 2025.
61. Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
62. Zhou Wang and Alan C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.
63. Qian Du, Nicholas H. Younan, Roger King, and Vijay P. Shah. On the performance evaluation of pan-sharpening techniques. *IEEE Geoscience and Remote Sensing Letters*, 4(4):518–522, 2007.
64. Abdelhamid N. Fsihan, Jean-Baptiste Thomas, Jon Y. Hardeberg, and Pierre Gouton. Spectral reconstruction from rgb imagery: A potential option for infinite spectral data? *Sensors*, 24(11):3666, 2024.
65. Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, 26, 2013.
66. Ali Mosleh, Luxi Zhao, Atin Singh, Jaeduk Han, Abhijith Punnappurath, Marcus A Brubaker, Jihwan Choe, and Michael S Brown. Non-parametric sensor noise modeling and synthesis. In *European Conference on Computer Vision*, pages 73–89. Springer, 2024.
67. Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.
68. Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.
69. Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021.
70. Seoung Wug Oh, Michael S Brown, Marc Pollefeys, and Seon Joo Kim. Do it yourself hyperspectral imaging with everyday digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2461–2469, 2016.

Hyperspectral reconstruction from RGB: review

D. M. Vladimirov, D. A. Reutsky, E. I. Ershov

Hyperspectral images (HSIs) contain, in each pixel, the distribution of light intensity across wavelengths. Due to the high cost of hyperspectral imaging equipment, researchers are developing methods to estimate the spectral of objects in the scene from compressed or low-dimensional spectral representations. This estimation process is commonly referred to as hyperspectral reconstruction (HR).

A widely studied approach is the reconstruction of HSIs from available RGB images of the scene. In this work, we present a review and classification of the most significant methods in this field, along with commonly used datasets and evaluation metrics for reconstruction quality assessment.

We conduct a computational experiments by retraining and testing all methods on the ARAD 1K dataset. Since the literature lacks a unified problem formulation for HR and does not specify the nature of the three-channel input images (RAW or JPEG), we perform experiments for both formats.

The best achieved PSNR values are 32.0 dB and 33.5 dB for reconstruction from JPEG and RAW images, respectively. When the test dataset is replaced with CAVE, the reconstruction quality of all methods degrades by at least a factor of two, indicating that the limited scene diversity in ARAD 1K restricts the generalization ability of HR models.

Finally, we discuss current challenges in hyperspectral reconstruction and outline potential directions for future research.

KEYWORDS: hyperspectral reconstruction, spectra, neural networks, transformers