

Оценка качества изображений с учётом моделей восприятия цветовых различий

Д. Э. Абрамов^{*,**}, С. А. Корчагин^{*,**}, Е. Е. Зайченкова^{*,**}, Е. И. Ершов^{*,**}

^{*}Институт проблем передачи информации им. А. А. Харкевича РАН, 127051, г. Москва, Б. Каретный пер., д.19, стр.1

^{**}Московский физико-технический институт (НИУ), 141701, Россия, Московская область, г. Долгопрудный, Институтский пер., 9

Поступила в редколлегию 01.07.2025 г. Принята 25.07.2025 г.

Аннотация—Наличие шума на изображениях является одной из наиболее значимых проблем в современной вычислительной фотографии, что объясняет безостановочное и бурное развитие данной области. Одним из наиболее сложных вопросов в данной области является поиск наилучшего метода оценки качества (МОК) изображения. Сегодня используются как аналитические МОКи, например, PSNR и SSIM, так и нейросетевые, например LPIPS и CLIP-IQA. В настоящий момент, что интересно, авторы не обнаружили МОК, которые учитывали бы знания об особенностях восприятия цвета стандартного наблюдателя CIE XYZ. В данной работе впервые показано, что учет такого знания с помощью формулы цветового различия CIEDE2000 и евклидова расстояния в пространстве CIELUV позволяет улучшить точность оценки качества изображения. В данной работе предлагается новый МОК на основе объединения классических и цветовых формул оценки различия между изображениями. Эксперименты на новом наборе данных показывают увеличение корреляции с результатами опросов испытуемых вплоть до 20%, подтверждая высокий потенциал развития методов в данном направлении.

КЛЮЧЕВЫЕ СЛОВА: оценка качества изображений, шумоподавление, цветовосприятие человека, средневзвешенная оценка мнений, цветовое пространство.

DOI: 10.53921/18195822_2025_25_2_114

1. ВВЕДЕНИЕ

Шумоподавление является ключевым этапом в процессе формирования изображений. На большинстве цифровых фотографий присутствуют шумы, обусловленные как физическими ограничениями сенсоров, так и неблагоприятными условиями съёмки. Во многих прикладных областях – таких как медицинская визуализация [28], спутниковая съёмка [29] и цифровая фотография – требуется получение изображений высокого качества, что делает задачу подавления шумов, искажений и артефактов особенно значимой.

Для построения эффективных алгоритмов шумоподавления необходимо надёжно оценивать уровень и характеристики шума. С этой целью применяются МОКи, позволяющие количественно измерять степень искажения по сравнению с эталонным изображением. Наиболее широко используются показатели PSNR (Peak Signal-to-Noise Ratio) и SSIM [1] (Structural Similarity Index Measure), а в последнее время предлагается множество их модификаций, а также принципиально новые методы оценки.

Однако существующие метрики обладают рядом ограничений. В частности, PSNR плохо отражает особенности человеческого восприятия, а SSIM не учитывает цветовые искажения. Более того, визуально различающиеся изображения могут иметь схожие значения этих метрик. Для устранения этих недостатков были предложены усовершенствованные версии этих

метрик, а также абсолютно новые подходы. Но и они в основном охватывают лишь отдельные аспекты визуального восприятия, при этом цветовая информация часто либо игнорируется, либо обрабатывается некорректно, без опоры на принципы колориметрии.

Чтобы устранить ограничения метрик, в работе [10] было предложено использовать нейросетевой подход для объединения нескольких метрик, обучая модель на наборах значений различных метрик. Качество такой модели оценивалось по степени корреляции её предсказаний с субъективными оценками качества (СОК) с использованием коэффициентов SROCC, PLCC и Kendall's τ . Но и в этой работе упущено влияние цветовых изменений.

В настоящем исследовании предполагается, что учёт цветовых искажений способен существенно повысить корреляцию между предсказанными метриками и субъективным восприятием. Для этого предлагается использовать показатели цветового различия CIEDE2000 и евклидово расстояние в пространстве CIELUV, которые рассчитываются после преобразования изображений в соответствующие цветовые координаты. Однако большинство используемых наборов данных не содержит информации о реальном цветовом пространстве изображений. В связи с этим в данной работе применяется гиперспектральный набор данных, обеспечивающий возможность корректного преобразования в необходимые пространства.

Развивая идеи, изложенные в [10], в данной работе предложена новая архитектура нейросети и метод формирования наборов метрик. Кроме того, для обучения использовалась также модель опорных векторов (SVM), результаты которой сопоставлены с нейросетевым подходом. Предложен новый набор данных, учитывающий цветовую составляющую. Проведённые эксперименты показали, что включение цветочувствительных метрик в обучающий набор значительно улучшает согласованность метрик с субъективной оценкой качества изображений. Показатели коэффициентов корреляции улучшились на 5–20%. Наилучшие достигнутые значения SROCC, PLCC и коэффициента Кендалла составили 0,913, 0,915 и 0,736 для модели SVM и 0,883, 0,899 и 0,701 — для нейронной сети соответственно. Схематическое представление предлагаемого алгоритма приведено на рисунке 1.

В данной работе представлен новый подход к учёту цветовой информации при оценке уровня шума на изображениях. В отличие от существующих методов, включая как классические метрики, так и модели, основанные на машинном обучении, предлагаемый подход направлен на выявление и количественную оценку цветовых искажений. Это становится возможным благодаря использованию гиперспектрального набора данных, что позволяет достичь более высокой согласованности с визуальным восприятием.

2. ОБЗОР МЕТРИК ШУМОПОДАВЛЕНИЯ

Как упоминалось выше, в настоящее время существует большое количество различных метрик шумоподавления. Рассмотрим некоторые из них.

PSNR (Peak Signal-to-Noise Ratio) — простая метрика, основанная на среднеквадратичной ошибке (MSE), выражаемой в децибелах. Она получила широкое распространение благодаря низкой вычислительной сложности. Однако PSNR слабо коррелирует с субъективным восприятием качества изображения, поскольку не учитывает структурные особенности изображений. PSNR-HVS [11], PSNR-HVS-M [12], PSNR-NA [13], PSNR-NMA [13] — усовершенствованные версии PSNR, учитывающие чувствительность зрения к различным типам искажений. Они снижают влияние незаметных артефактов и увеличивают вклад значимых областей. Эти метрики точнее коррелируют с визуальным восприятием, особенно при неравномерных искажениях, но при этом сохраняют простую структуру расчёта.

SSIM [1] (Structural Similarity Index) — оценивает качество на основе сходства структуры, яркости и контраста. Лучше отражает восприятие локальных искажений по сравнению с PSNR. Однако чувствительна к масштабным изменениям и может быть менее точна при силь-

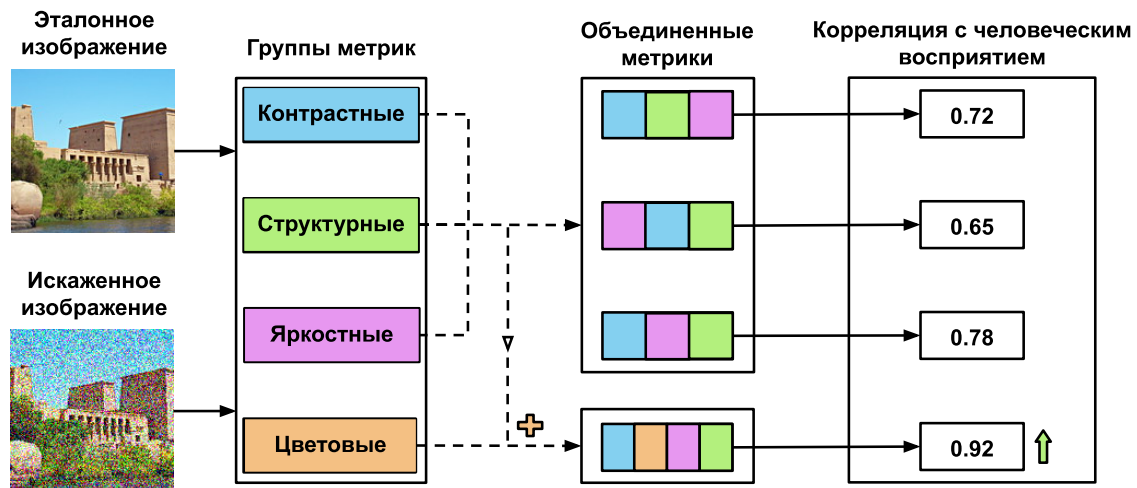


Рис. 1. Общая идея данной работы заключается в использовании сразу нескольких метрик разного вида, объединяя их в единую нейросетевую метрику. Использование цветовых формул в наборах позволяет значительно увеличить корреляцию нейросетевой метрики с человеческим восприятием.

ных искажениях. ESSIM [7] — модификация SSIM с акцентом на сравнение краевых структур. Обеспечивает лучшую чувствительность к искажениям контуров, но требует более сложной обработки границ. UIQI [9] (Universal Image Quality Index) — ранняя версия SSIM, анализирующая локальное сходство по яркости, контрасту и структуре. Обеспечивает хорошую интерпретируемость и низкую сложность, но менее точна при сложных искажениях.

FSIM [6] (Feature Similarity Index) — использует фазовую и градиентную информацию, подражая восприятию человеческого глаза. Отличается высокой корреляцией с субъективным качеством, особенно на сложных текстурах. Недостаток — высокая вычислительная стоимость. VSI [2] (Visual Saliency-Induced Index) — сочетает структурные и цветовые признаки с картой визуальной значимости. Модель визуального внимания выделяет наиболее важные участки сцены, что улучшает соответствие субъективным оценкам. Однако метрика чувствительна к точности этой модели. Метрика MDSI [3] (Mean Deviation Similarity Index) предназначена для оценки визуального качества изображений с учётом структурных, градиентных и цветовых различий. Она сочетает яркостное и цветовое сходство с использованием отклонений от идеального совпадения, что позволяет точно учитывать как глобальные, так и локальные искажения. Благодаря взвешенному объединению нескольких компонентов MDSI обеспечивает высокую корреляцию с субъективными оценками при разнообразных типах искажений. QILV [8] — основана на локальной дисперсии яркости, отражающей степень текстурной детализации. Эффективна в выявлении потерь визуально значимых деталей, особенно в вариативных областях. Может быть менее чувствительной к искажениям в гладких зонах. GMSD [5] (Gradient Magnitude Similarity Deviation) — рассчитывает отклонения градиентных величин по изображению. Отличается высокой чувствительностью к визуальным дефектам при низкой вычислительной сложности. Подходит для быстрой оценки, но не учитывает цветовую информацию.

На момент публикации данной работы уже существует ряд исследований, посвящённых применению методов машинного обучения и нейронных сетей для оценки качества изображений. Наиболее распространённым подходом является использование сверточных нейронных сетей (CNN). Так, в работе [14] CNN применяется для извлечения признаков из искажённых и эталонных фрагментов изображения с последующим прогнозированием воспринимаемого качества; итоговая оценка качества изображения формируется путём агрегации локальных предсказаний. В исследовании [15] предлагается архитектура CNN, ориентированная на анализ цветовой информации в различных цветовых пространствах: признаки извлекаются из цветных изображений и используются для предсказания субъективной оценки качества, при этом осуществляется преобразование изображений в дополнительные цветовые представления. Некоторые методы, например [16], допускают использование как в режиме с эталоном (FR, Full Reference), так и без него (NR, No Reference), в зависимости от доступности оригинального изображения. Отдельного внимания заслуживает работа [17], в которой для объединения различных метрик качества применяется классический подход машинного обучения. С использованием метода опорных векторов (SVM) формируется обобщённая оценка качества, обеспечивающая высокую корреляцию с субъективными оценками. При этом выбор входных метрик осуществляется автоматически с учётом их свойств и области применения; дополнительно учитываются цветовые различия и применяются преобразования в альтернативные цветовые пространства.

3. ОБЗОР НАБОРОВ ДАННЫХ

Во всех вышеупомянутых исследованиях в качестве основы для обучения и оценки моделей используются широко распространённые наборы данных, такие как LIVE [20], CSIQ [21], TID2013 [22] и TID2008 [23].

Набор данных LIVE, созданный в 2006 году, включает 779 изображений с различными сценами (портреты, пейзажи, макро- и широкоугольные снимки), подвергнутых пяти типам искажений: JPEG2000 и JPEG-сжатие, аддитивный белый шум, размытие по Гауссу и затухание по модели Рэля. CSIQ содержит 866 искажённых изображений, полученных из 30 эталонных, с применением таких искажений, как JPEG-сжатие, гауссово размытие, белый и розовый шум, а также изменение контрастности. Набор TID2013 включает 25 эталонных изображений и 3000 искажённых, охватывая 24 типа искажений с пятью уровнями искажения каждого типа.

Однако все эти наборы данных имеют важное ограничение с точки зрения задач, связанных с анализом цветовых искажений. Изображения в них были получены с использованием плёночных фотоаппаратов, а затем прошли несколько этапов обработки — проявление, сканирование, оцифровка, — в результате чего точная информация о цветовом пространстве была утрачена. Это существенно ограничивает возможность корректного анализа цветовых искажений и применения колориметрических метрик. В связи с этим для настоящей работы был собран собственный набор данных, удовлетворяющий необходимым требованиям и позволяющий точно учитывать цветовые характеристики изображений. Его структура и свойства подробно описаны в разделе 4.

4. НАБОРЫ МЕТРИК

Для создания наборов использовались следующие метрики: PSNR, SSIM [1], FSIM [6], VSI [2], GMSD [5], MS-SSIM [24], VIF [25], LPIPS [26], CIEDE2000 в цветовом пространстве CIELAB [19], L2 в цветовом пространстве CIELUV [27]. Опишем кратко каждую из них.

Широко используемая метрика **PSNR** задается формулой:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{(R_{\max})^2}{\text{MSE}} \right),$$

где:

- R_{\max} — максимальное значение пикселя изображения. Например, для 8-битных изображений $R_{\max} = 255$,
- MSE — среднеквадратическая ошибка между оригинальным изображением I_{orig} и искажённым изображением I_{dist} .

Среднеквадратическая ошибка (MSE) вычисляется по формуле:

$$\text{MSE} = \frac{1}{N \cdot M} \sum_{i=1}^N \sum_{j=1}^M (I_{\text{orig}}(i, j) - I_{\text{dist}}(i, j))^2,$$

где:

- N и M — размеры изображения (в пикселях),
- $I_{\text{orig}}(i, j)$ и $I_{\text{dist}}(i, j)$ — значения пикселей на позиции (i, j) для исходного и искажённого изображений соответственно.

Высокие значения PSNR указывают на высокое качество изображения (меньшее количество искажений), а низкие значения свидетельствуют о большем уровне шума или искажений.

Метрика **SSIM** [1] учитывает изменение яркости, контраста и структуры. Её можно вычислить по следующему алгоритму. Изображения I_{orig} и I_{dist} разбиваются на маленькие окна x и y (например, 8x8 или 11x11 пикселей). Для каждого окна x и y вычисляются следующие параметры:

$$\begin{aligned} \mu_x &= \frac{1}{N} \sum_{i=1}^N x_i, & \mu_y &= \frac{1}{N} \sum_{i=1}^N y_i \\ \sigma_x &= \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)^2}, & \sigma_y &= \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \mu_y)^2} \\ \rho_{xy} &= \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \end{aligned}$$

Затем для каждого окна вычисляются компоненты яркости, контраста и структуры:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(x, y) = \frac{\rho_{xy} + C_3}{\sigma_x\sigma_y + C_3},$$

где C_1, C_2, C_3 — малые постоянные величины для предотвращения деления на ноль (например, $C_1 = 10^{-4}$, $C_2 = 10^{-4}$, $C_3 = C_2/2$). Для каждого окна x и y вычисляется SSIM:

$$\text{SSIM}(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y)$$

Общее значение SSIM для изображения вычисляется как среднее значение SSIM по всем окнам:

$$\text{SSIM}_{\text{image}} = \frac{1}{N_{\text{windows}}} \sum_{i=1}^{N_{\text{windows}}} \text{SSIM}(x_i, y_i),$$

где $N_windows$ - число окон.

Метрика **FSIM** [6] оценивает качество изображения, сравнивая ключевые визуальные признаки — фазу PC и градиенты GM . Для каждого пикселя вычисляются карты сходства:

$$S_{PC}(x) = \frac{2 PC_I(x) PC_{I'}(x) + T_1}{PC_I(x)^2 + PC_{I'}(x)^2 + T_1}, \quad S_{GM}(x) = \frac{2 GM_I(x) GM_{I'}(x) + T_2}{GM_I(x)^2 + GM_{I'}(x)^2 + T_2}$$

Общая карта сходства:

$$S_L(x) = S_{PC}(x) \cdot S_{GM}(x)$$

Итоговый индекс вычисляется как взвешенное среднее:

$$FSIM = \frac{\sum_x S_L(x) \cdot W(x)}{\sum_x W(x)},$$

где $W(x)$ — весовая карта значимости, отражающая вклад каждого пикселя в итоговую оценку.

Метрика **VSI** [2] оценивает качество изображения, учитывая визуальную значимость различных областей на основе карты внимания (saliency map) $S(x)$, структурного сходства $S_{str}(x)$ и цветового сходства $S_{col}(x)$. Общая карта сходства вычисляется как произведение:

$$S_{VSI}(x) = S_{str}(x) \cdot S_{col}(x)$$

Итоговый индекс **VSI** рассчитывается как взвешенное среднее по всему изображению:

$$VSI = \frac{\sum_x S_{VSI}(x) \cdot S(x)}{\sum_x S(x)},$$

где $S(x)$ — карта визуальной значимости, выделяющая наиболее заметные для наблюдателя области. Таким образом, **VSI** придаёт больший вес визуально важным участкам, улучшая согласованность с субъективными оценками качества.

Метрика **GMSD** [5] оценивает качество изображения на основе сравнения градиентных величин эталонного и искажённого изображений, отражающих локальные изменения яркости и структуру объектов. Для каждого пикселя вычисляется индекс сходства градиентов:

$$GMS(x) = \frac{2 \cdot GM_I(x) \cdot GM_{I'}(x) + T}{GM_I(x)^2 + GM_{I'}(x)^2 + T},$$

где $GM_I(x)$ и $GM_{I'}(x)$ — градиентные величины эталонного и искажённого изображений в точке x , T — константа для численной устойчивости. Итоговый индекс **GMSD** рассчитывается как стандартное отклонение карты сходства градиентов:

$$GMSD = \sqrt{\frac{1}{N} \sum_{x=1}^N (GMS(x) - \overline{GMS})^2},$$

где N — число пикселей, а \overline{GMS} — среднее значение карты сходства градиентов. Такой подход позволяет выявлять локальные искажения, существенно влияющие на визуальное качество, обеспечивая высокую чувствительность к заметным дефектам.

Метрика **MS-SSIM** [24] оценивает качество изображения, анализируя сходство между эталонным и искажённым изображениями на нескольких масштабах. На каждом уровне j вычисляются компоненты сходства:

$$l_j = \frac{2\mu_x^{(j)}\mu_y^{(j)} + C_1}{(\mu_x^{(j)})^2 + (\mu_y^{(j)})^2 + C_1}, \quad c_j = \frac{2\sigma_x^{(j)}\sigma_y^{(j)} + C_2}{(\sigma_x^{(j)})^2 + (\sigma_y^{(j)})^2 + C_2}, \quad s_j = \frac{\sigma_{xy}^{(j)} + C_3}{\sigma_x^{(j)}\sigma_y^{(j)} + C_3},$$

где $\mu_x^{(j)}, \mu_y^{(j)}$ — средние значения, $\sigma_x^{(j)}, \sigma_y^{(j)}$ — стандартные отклонения, а $\sigma_{xy}^{(j)}$ — ковариация локальных окон эталонного и искажённого изображений на уровне j . Итоговый индекс MS-SSIM вычисляется как:

$$MS-SSIM = l_M^{\alpha_M} \cdot \prod_{j=1}^M c_j^{\beta_j} s_j^{\gamma_j},$$

где M — число уровней масштабирования, а $\alpha_j, \beta_j, \gamma_j$ — веса, задающие вклад каждого компонента на соответствующем уровне. Такой многоуровневый подход позволяет учитывать качество изображения на разных деталях и масштабах, улучшая корреляцию с субъективной оценкой.

Метрика **VIF** [25] (Visual Information Fidelity) оценивает качество изображения через количество визуальной информации, сохранённой в искажённом изображении относительно оригинала. Для этого изображения разбиваются на несколько уровней детализации (подполос), каждый из которых содержит разные детали изображения. Пусть на уровне k эталонное изображение представлено коэффициентами C_k , а искажённое — D_k . Модель описывается как:

$$D_k = g_k C_k + v_k,$$

где g_k — коэффициент усиления, а v_k — аддитивный шум. Визуальная информация в каждой подполосе оценивается через взаимную информацию:

$$I(C_k; F_k) = \frac{1}{2} \sum_i \log_2 \left(1 + \frac{g_{k,i}^2 \sigma_{C_{k,i}}^2}{\sigma_{v_{k,i}}^2} \right),$$

где $\sigma_{C_{k,i}}^2$ и $\sigma_{v_{k,i}}^2$ — дисперсии сигнала и шума для коэффициента i на уровне k . Итоговое значение VIF вычисляется как отношение суммарной информации, переданной искажённым изображением, к информации эталонного:

$$VIF = \frac{\sum_k I(C_k; F_k)}{\sum_k I(C_k; E_k)},$$

где E_k — шум в эталонном изображении. Таким образом, VIF отражает, какая часть визуальной информации оригинала сохраняется в искажённом изображении, что хорошо коррелирует с субъективным восприятием качества.

Метрика **LPIPS** [26] вычисляет сходство изображений на основе расстояния между признаковыми представлениями, извлечёнными из предобученной сверточной нейронной сети. Пусть $F_i(x)$ и $F_i(y)$ — активации i -го слоя сети для эталонного изображения x и искажённого y , соответственно. Сначала эти активации нормируются по каналам:

$$\hat{F}_i(z) = \frac{F_i(z)}{\|F_i(z)\|_2}$$

Далее вычисляется расстояние на слое i :

$$d_i(x, y) = \frac{1}{H_i W_i} \sum_{h=1}^{H_i} \sum_{w=1}^{W_i} \left\| w_i \odot \left(\hat{F}_i(x)_{h,w} - \hat{F}_i(y)_{h,w} \right) \right\|_2^2,$$

где w_i — обучаемый вектор весов по каналам, H_i, W_i — высота и ширина карты признаков. Итоговая метрика LPIPS вычисляется как сумма взвешенных расстояний по выбранным слоям:

$$LPIPS(x, y) = \sum_i d_i(x, y).$$

Такой подход позволяет учитывать высокоуровневые визуальные различия, близкие к восприятию человека.

Особое место среди рассматриваемых нами метрик занимают специализированные формулы цветового различия, такие как метрика L2 в цветовом пространстве CIELUV и CIEDE2000 [19] в пространстве CIELAB.

Метрика **L2** вычисляется в цветовом пространстве CIELUV [27] по формуле:

$$L2 = \sqrt{(L_1 - L_2)^2 + (u_1 - u_2)^2 + (v_1 - v_2)^2}, \quad (1)$$

где (L_1, u_1, v_1) и (L_2, u_2, v_2) – два цвета в пространстве CIELUV. Пространство CIELUV является перцептивно униформным, то есть сконструировано таким образом, чтобы евклидово расстояние между двумя точками в этом пространстве приблизительно соответствовало субъективному восприятию цветовой разницы. Благодаря этому значения L2-метрики в пространстве CIELUV хорошо коррелируют с визуально воспринимаемыми различиями между цветами. В отличие от метрик, ориентированных на яркость, контраст или структуру, L2 в CIELUV позволяет выявлять и количественно оценивать именно цветовые искажения, что делает её полезным дополнением к стандартным показателям качества изображений.

Формула **CIEDE2000** [19] оценивает цветовое различие между двумя цветами в пространстве CIELAB с учётом особенностей восприятия человеческого глаза. Алгоритм включает следующие этапы:

1. Преобразование цветов в параметры L^* (яркость), a^* , b^* (цветовые компоненты).
2. Вычисление скорректированных различий яркости $\Delta L'$, насыщенности $\Delta C'$ и оттенка $\Delta H'$, с учётом нелинейностей восприятия.
3. Учёт взаимодействия насыщенности и оттенка, а также применение весовых функций S_L, S_C, S_H и параметров масштабирования k_L, k_C, k_H .
4. Расчёт итогового цветового расстояния по формуле:

$$\Delta E_{2000} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \cdot \frac{\Delta C'}{k_C S_C} \cdot \frac{\Delta H'}{k_H S_H}},$$

где R_T – термин, учитывающий взаимодействие насыщенности и оттенка.

Метрика CIEDE2000 обеспечивает более точное соответствие восприятию человеком цветовых различий и широко используется в стандартах цветовых измерений.

В отличие от общих цветовых метрик, эти методы сосредоточены исключительно на оценке цветовых изменений, не учитывая структурные или яркостные характеристики изображения. Например, в метриках FSIM, VSI и GMSD цветовые отклонения учитываются поверхностно и с меньшей точностью, при этом данные методы не опираются на принципы колориметрии. Предполагается, что включение специализированных формул цветового различия в состав наборов позволит повысить точность оценки цветовых искажений.

В данной работе выдвигается гипотеза, что обучение нейросетевой метрики на наборах, включающих метрики, охватывающие широкий спектр свойств изображений, важных для восприятия человеком, позволит добиться более высокой корреляции с субъективной оценкой качества. Это превосходит подходы, основанные на использовании наборов метрик, учитывающих только однотипные параметры изображения. Для реализации этой идеи перечисленные метрики были разделены на несколько групп (см. таблицу 1). Каждая группа отражает определённый аспект изображения, который учитывается при вычислении оценки качества. Такой

Таблица 1. Разбиение метрик по группам в соответствии с параметрами изображений.

Цветовые	Структурные	Яркостные	Контрастные
CIEDE2000 [19]	SSIM [1]	SSIM [1]	SSIM [1]
L2	MS-SSIM [24]	MS-SSIM [24]	MS-SSIM [24]
	FSIM [6]	PSNR	VSI [2]
	VSI [2]	LPIPS [26]	GMSD [5]
	GMSD [5]		FSIM [6]
	VIF [25]		VIF [25]
	LPIPS [26]		LPIPS [26]

подход позволяет обеспечить более комплексное представление характеристик изображения при обучении модели.

Цветовые формулы оценивают искажения, возникающие в цветовом восприятии изображения, фиксируя изменения в цветовых компонентах между оригиналом и искажённой версией. Такие метрики, как правило, основаны на представлении изображений в перцептивно равномерных цветовых пространствах (например, CIELAB [19], CIELUV [27]) и позволяют количественно описывать различия в оттенках, насыщенности и цветовой температуре. Структурные метрики направлены на анализ пространственных зависимостей и геометрических свойств изображения. Они измеряют, насколько сохранены важные визуальные паттерны, текстура, границы объектов и другие структурные элементы. Яркостные метрики оценивают разницу в яркостных уровнях между оригиналом и искажённым изображением. Они анализируют абсолютные или относительные отклонения яркости. Контрастные метрики чувствительны к изменениям в различии между светлыми и тёмными областями изображения. Они оценивают, насколько хорошо различаются детали при изменениях освещённости и насколько искажения повлияли на визуальное разделение объектов.

В рамках проведённой классификации такие метрики, как FSIM, GMSD, LPIPS и другие, не были отнесены к цветовым, поскольку, как отмечалось ранее, они не измеряют цветовые искажения напрямую и не опираются на принципы колориметрии. Их потенциальная чувствительность к цветовым изменениям имеет лишь косвенный характер и связана с реакцией на цветовые искажения в процессе общего анализа, без явного моделирования механизмов цветового восприятия человека.

5. НАБОР ДАННЫХ

Как упоминалось выше, в конце раздела 2, общедоступные наборы данных, используемые в подобных задачах, не подходят для оценки изменения цвета, т.к. неизвестно, в каком цветовом пространстве находятся изображения из этих наборов.

Для решения поставленной задачи при обучении моделей машинного обучения и вычислении метрик качества изображений в данной работе использовался набор данных, основанный на изображениях, полученных с гиперспектральной камеры.

Гиперспектральные изображения представляют собой трёхмерные массивы, в которых каждый пиксель содержит спектр отражённого света, измеренный с высокой спектральной дискретизацией. Такой формат данных обеспечивает гораздо более полное описание визуальных свойств сцены по сравнению с обычными RGB-изображениями и позволяет моделировать цветопередачу с учётом физических характеристик освещения и чувствительности наблюдателя.

Наличие спектральной информации в каждом пикселе делает возможным перевод таких изображений в колориметрические цветовые пространства, что, в свою очередь, позволяет

рассчитывать цветовые формулы, чувствительные к различиям, неуловимым для традиционных метрик.

В качестве исходных гиперспектральных данных были использованы 70 изображений из набора ARAD_1K [32], содержащего 1000 гиперспектральных сцен. Изображения были получены с использованием мобильной гиперспектральной камеры Specim IQ с исходным спектральным диапазоном от 400 до 1000 нм. В настоящей работе использовалась версия набора, содержащая 31 спектральный канал, равномерно распределённый в диапазоне 400–700 нм с шагом около 10 нм.

Выполнялось преобразование гиперспектральных данных в цветовое пространство CIE XYZ [34]. Отражательные спектры пикселей $R(\lambda)$ предварительно корректировались по освещению путём поэлементного умножения на спектральную характеристику стандартного источника освещения D50 $E(\lambda)$, формируя результирующий спектр:

$$S(\lambda) = R(\lambda) \cdot E(\lambda).$$

Затем для каждого пикселя рассчитывались координаты CIE XYZ путём дискретного интегрирования с использованием функций цветового соответствия стандартного наблюдателя CIE 1931 (2°):

$$X = \sum_{\lambda} S(\lambda) \cdot \bar{x}(\lambda) \cdot \Delta\lambda, \quad Y = \sum_{\lambda} S(\lambda) \cdot \bar{y}(\lambda) \cdot \Delta\lambda, \quad Z = \sum_{\lambda} S(\lambda) \cdot \bar{z}(\lambda) \cdot \Delta\lambda,$$

где $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ — функции цветового соответствия, $\Delta\lambda$ — шаг дискретизации (10 нм).

Для генерации обучающих примеров к каждому полученному изображению добавлялся цветовой шум. Это позволяло продемонстрировать ограничения традиционных подходов и подчеркнуть необходимость использования цветочувствительных методов оценки качества изображений. Это сделано для проверки гипотезы о том, что человеческая зрительная система сильно чувствительна к изменениям цветов, которые классические метрики не улавливают. К каждому из цветовых каналов изображения ($c = r, g, b$) применялся Гауссовский шум с определенным весом w_c :

$$NoisyImage_c = img_c + w_c \cdot N(0; 0, 046) \quad (2)$$

Шум применялся к изображениям в пространстве CIE XYZ. Были выбраны следующие наборы весов:

1. $w_r, w_g, w_b = \{1; 0; 0\}; \{0; 1; 0\}; \{0; 0; 1\}$.
2. $w_r, w_g, w_b = \{0, 1; 0, 3; 0, 6\}; \{0, 6; 0, 3; 0, 1\}; \{0, 1; 0, 6; 0, 3\}$.

Таким образом, на основе 70 начальных изображений было получено 420 зашумленных: к каждому изображению применялся шум с одной из шести указанных выше комбинацией весовых коэффициентов. Полученный набор данных был разделен на обучающую, валидационную и тестовую выборки в соотношении 8:1:1. Для всех изображений были посчитаны значения каждой метрики, которые будут выступать а качестве признаков при обучении.

В качестве целевой переменной, которую хотим предсказывать, будет использоваться метрика СОК – средневзвешенная оценка мнений (Mean Opinion Score). Для подсчета значений СОК для каждого изображения использовался сервис Яндекс.Задания: в каждом задании испытуемому показывали пару изображений и просили выбрать менее зашумленное на его взгляд. Если испытуемому казалось, что уровень шума одинаковый, то выбирали ответ «Изображения одинаковые», такие ответы не учитывались при подсчете количества выборов изображения. Всего на каждое сравнение (на каждую пару изображений) было получено по 15 ответов. СОК для изображения считался как отношение числа выборов его в качестве менее шумного к общему числу его показов и по аналогии с [10] нормировался в диапазон от 0 до 9.

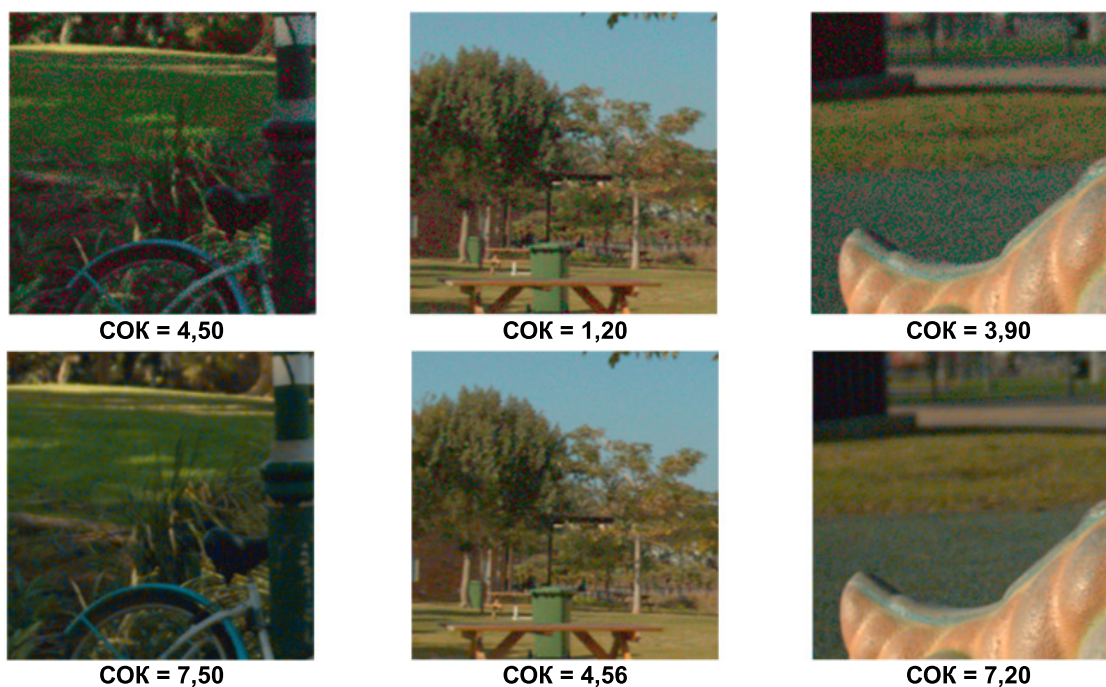


Рис. 2. Примеры изображений из набора данных.

$$\text{СОК} = 9 \times \frac{\text{Количество выборов изображения}}{\text{Количество показов изображения}}$$

Сгенерированный набор данных показал, что человеческое зрение действительно очень чувствительно к цветовому шуму. На рисунке 2 приведены примеры изображений из набора данных.

Исследование [33] показывает, что такой метод сбора СОК может быть использован вместо лабораторных исследований с некоторой степенью точности.

6. НЕЙРОСЕТЕВАЯ МЕТРИКА И SVM

Для обучения нейронной сети на её вход подаётся вектор, содержащий значения метрик, вычисленных между текущим изображением из набора и соответствующим эталонным изображением. Длина этого вектора соответствует числу используемых метрик, в связи с чем для каждого набора метрик применяется индивидуальная архитектура сети с соответствующим числом входных нейронов.

По этим значениям нейронная сеть учится предсказывать одну величину — значение СОК, то есть выходными данными сети является одно число. Полученное значение является оценкой качества изображения, которую дает нейросетевая метрика.

Для обучения сети использовалась следующая архитектура: два скрытых слоя, каждый из которых содержит по 16 нейронов с функцией активации ELU (Exponential Linear Unit), за первым слоем с активацией следуют Dropout с вероятностью зануления 0.2 для регуляризации. Оптимизация выполняется с помощью Adam (learning rate = 0.00025, weight_decay = 0), функция потерь – L1 Loss (средняя абсолютная ошибка – MAE). Выходной слой состоит из одного нейрона и используется для получения финального предсказания.

Кроме того были проведены эксперименты с методом опорных векторов (SVM) для задачи регрессии. Для повышения точности предсказания и стабильности обобщения применялась стандартная предварительная обработка данных с использованием стандартизации признаков (StandardScaler). Подбор гиперпараметров (ϵ , C и тип ядра: RBF или линейное) выполнялся с использованием перекрёстной проверки (5-fold cross-validation) и процедуры GridSearchCV, где метрикой качества служила отрицательная среднеквадратичная ошибка.

На основе таблицы 1 были сформированы наборы метрик. Объединение нескольких метрик выполнялось так, чтобы в совокупности они охватывали все параметры, перечисленные в таблице 1. Таким образом, нейросетевая метрика, использующая подобный набор, будет учитывать большее количество разнообразных характеристик изображения, что позволяет максимально приблизить ее оценку к оценке зрительной системы человека. Составленные наборы приведены в таблице 2.

Таблица 2. Наборы метрик, используемые для обучения моделей машинного обучения.

№ набора	Метрики в наборе
1	PSNR, SSIM
2	PSNR, SSIM, CIEDE2000, L2
3	PSNR, FSIM, GMSD
4	PSNR, FSIM, GMSD, CIEDE2000, L2
5	FSIM, VSI, GMSD
6	FSIM, VSI, GMSD, CIEDE2000, L2
7	MS-SSIM, VSI, LPIPS
8	MS-SSIM, VSI, LPIPS, CIEDE2000, L2
9	VIF, GMSD, LPIPS
10	VIF, GMSD, LPIPS, CIEDE2000, L2
11	PSNR, SSIM, MS-SSIM
12	PSNR, SSIM, MS-SSIM, CIEDE2000, L2

Наборы метрик формировались следующим образом: сначала создавался набор, включающий только стандартные (нецветовые) метрики, а затем — аналогичный набор, дополненный цветовыми формулами. При помощи этих наборов будет учтен вклад узконаправленных метрик цветовых различий с переходом в цветовые пространства. Схематичное описание экспериментов приведено на рис. 3.

7. МЕТРИКИ КОРРЕЛЯЦИИ

Для контроля качества итоговой оценки используются три коэффициента корреляции: SROCC (Spearman's Rank-Order Correlation Coefficient), PLCC (Pearson Linear Correlation Coefficient) и Kendall's τ (коэффициент Кендалла). Это статистические меры, которые часто применяются в задачах оценки качества изображений, видео, звука и других сигналов, чтобы сравнить автоматическую метрику с субъективными оценками, например, СОК (Средневзвешенная Оценка Мнений). Они показывают, насколько хорошо метрика «предсказывает» или «согласуется» с человеческим восприятием.

SROCC измеряет степень согласованности порядка между двумя наборами данных. Он основан на сравнении рангов, а не самих значений, что делает его устойчивым к нелинейным преобразованиям. Значение коэффициента лежит в диапазоне от -1 до 1 , где 1 соответствует

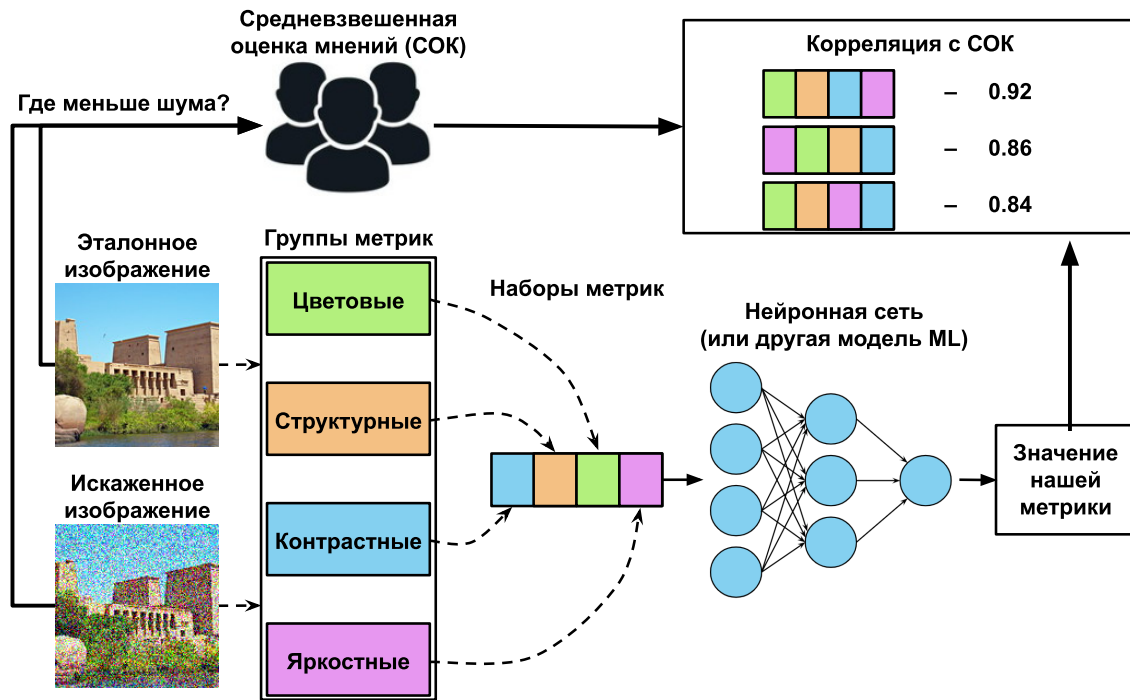


Рис. 3. Схема проведения экспериментов.

идеальному совпадению рангов. Вычисляется по формуле:

$$\text{SROCC} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

где d_i — разность рангов, а n — число наблюдений.

PLCC отражает степень линейной зависимости между предсказанными значениями и эталонными оценками. В отличие от SROCC, этот коэффициент чувствителен к масштабированию и смещению. Его значения также лежат в диапазоне от -1 до 1 . Он рассчитывается как:

$$\text{PLCC} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}},$$

где x_i и y_i — предсказанные и истинные значения, \bar{x} и \bar{y} — их средние.

Коэффициент **Kendall's τ** также измеряет порядковую согласованность между двумя переменными, но делает это путём анализа пар наблюдений. Он рассчитывает разность между количеством согласованных и несогласованных пар:

$$\tau = \frac{C - D}{\frac{1}{2}n(n - 1)},$$

где C — число конкордантных пар, D — дискордантных.

SROCC и Kendall's τ оценивают ранговую (порядковую) корреляцию и не чувствительны к линейности. PLCC, напротив, характеризует степень линейного соответствия между метриками. Совместное использование всех трёх коэффициентов позволяет всесторонне оценить поведение предлагаемой метрики качества по сравнению с субъективными данными.

8. ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ

На каждом из перечисленных выше наборов была рассчитана нейросетевая метрика и метод опорных векторов. Результаты данных экспериментов приведены в таблицах 3 и 4.

Таблица 3. Результаты работы **полносвязной нейронной сети** для задачи регрессии на наборах без и с использованием цветowych формул.

Набор метрик	SROCC	PLCC	Kendall
PSNR, SSIM	0,757	0,761	0,572
PSNR, SSIM + ΔE_{2000} , $L2_{Luv}$	0,784	0,814	0,605
PSNR, FSIM, GMSD	0,753	0,764	0,568
PSNR, FSIM, GMSD + ΔE_{2000} , $L2_{Luv}$	0,864	0,868	0,689
FSIM, VSI, GMSD	0,771	0,774	0,583
FSIM, VSI, GMSD + ΔE_{2000} , $L2_{Luv}$	0,778	0,783	0,603
MS-SSIM, VSI, LPIPS	0,653	0,697	0,477
MS-SSIM, VSI, LPIPS + ΔE_{2000} , $L2_{Luv}$	0,721	0,701	0,539
VIF, GMSD, LPIPS	0,729	0,714	0,550
VIF, GMSD, LPIPS + ΔE_{2000} , $L2_{Luv}$	0,775	0,804	0,602
PSNR, SSIM, MS-SSIM	0,743	0,749	0,550
PSNR, SSIM, MS-SSIM + ΔE_{2000} , $L2_{Luv}$	0,846	0,877	0,663
Полный набор метрик	0,844	0,832	0,652
Полный набор метрик + ΔE_{2000} , $L2_{Luv}$	0,883	0,899	0,701

Полносвязная нейронная сеть, реализованная согласно описанной выше архитектуре, демонстрирует существенно более высокие показатели точности (см. таблицу 3). При использовании набора метрик PSNR, FSIM и GMSD в комбинации с формулами цветового различия достигаются значения коэффициента ранговой корреляции Спирмена (SROCC) = 0,864 и коэффициента Кендалла (Kendall's τ) = 0,689. При применении метрик PSNR, SSIM и MS-SSIM совместно с цветовыми формулами наблюдается высокий уровень линейной зависимости: коэффициент линейной корреляции Пирсона (PLCC) составляет 0,877. Во всех проведённых экспериментах зафиксировано крайне малое значение p-value (порядка 3, 15×10^{-20}), что указывает на чрезвычайно низкую вероятность случайного получения наблюдаемых корреляционных связей при отсутствии действительной зависимости. Это позволяет с высокой степенью статистической достоверности утверждать наличие значимой взаимосвязи между анализируемыми переменными.

Кроме того, для данной модели отмечается, что включение формул цветового различия приводит к повышению точности предсказания в диапазоне от 3% до 15%.

Рассмотрим результаты метода опорных векторов (SVM), представленные в таблице 4. Данный подход демонстрирует качество предсказания, сопоставимое с нейросетевой моделью. Наилучшие значения метрик достигаются при использовании аналогичных комбинаций признаков. В частности, при применении метрик PSNR, FSIM и GMSD совместно с цветовыми формулами достигается коэффициент ранговой корреляции Спирмена (SROCC) = 0,835. Комбинация PSNR, SSIM, MS-SSIM и цветowych методов обеспечивает коэффициент линейной корреляции Пирсона (PLCC) = 0,854. Максимальное значение коэффициента Кендалла (Kendall's τ), равное 0,642, достигается при использовании набора VIF, GMSD и LPIPS в сочетании с формулами цветового различия.

Аналогично результатам, полученным с использованием нейронной сети, во всех экспериментах наблюдаются чрезвычайно низкие значения p-value (около 4, 28×10^{-15}), что подтверждает статистическую значимость выявленных зависимостей. Включение цветowych формул

Таблица 4. Результаты работы Метода Опорных Векторов (SVM) для задачи регрессии на наборах без и с использованием цветowych формул.

Набор метрик	SROCC	PLCC	Kendall	Лучшие гиперпараметры
PSNR, SSIM	0,664	0,683	0,478	C: 100, ϵ : 0,5
PSNR, SSIM + ΔE_{2000} , $L2_{Luv}$	0,779	0,802	0,560	C: 10, ϵ : 0,5
PSNR, FSIM, GMSD	0,705	0,723	0,529	C: 10, ϵ : 0,5
PSNR, FSIM, GMSD + ΔE_{2000} , $L2_{Luv}$	0,835	0,849	0,632	C: 10, ϵ : 0,1
FSIM, VSI, GMSD	0,682	0,680	0,512	C: 10, ϵ : 0,001
FSIM, VSI, GMSD + ΔE_{2000} , $L2_{Luv}$	0,794	0,767	0,614	C: 10, ϵ : 0,5
MS-SSIM, VSI, LPIPS	0,626	0,606	0,466	C: 1, ϵ : 0,5
MS-SSIM, VSI, LPIPS + ΔE_{2000} , $L2_{Luv}$	0,804	0,800	0,616	C: 10, ϵ : 0,5
VIF, GMSD, LPIPS	0,702	0,721	0,512	C: 1, ϵ : 0,5
VIF, GMSD, LPIPS + ΔE_{2000} , $L2_{Luv}$	0,828	0,823	0,642	C: 10, ϵ : 0,5
PSNR, SSIM, MS-SSIM	0,709	0,745	0,528	C: 10, ϵ : 0,01
PSNR, SSIM, MS-SSIM + ΔE_{2000} , $L2_{Luv}$	0,828	0,854	0,621	C: 100, ϵ : 0,5
Полный набор метрик	0,765	0,777	0,590	C: 10, ϵ : 0,01
Полный набор метрик + ΔE_{2000} , $L2_{Luv}$	0,913	0,915	0,736	C: 100, ϵ : 0,5

стабильно приводит к улучшению качества: в отдельных случаях прирост значений корреляционных коэффициентов достигает 18%.

В последнем столбце таблицы 4 представлены оптимальные значения гиперпараметров, при которых были достигнуты соответствующие показатели корреляции. Наилучшее значение гиперпараметра 'kernel', полученное во всех проведённых экспериментах, соответствует ядру типа 'rbf'.

Дополнительно к описанным экспериментам был сформирован объединённый набор, включающий все используемые метрики: PSNR, SSIM, MS-SSIM, FSIM, VSI, GMSD, VIF, LPIPS вместе с цветовыми CIEDE2000, L2. Обучение вышеупомянутых моделей на данном объединённом наборе позволило достичь максимальных значений коэффициентов корреляции, которые представлены в последних строках таблиц 3 и 4.

Таблица 5. Значения коэффициентов корреляции для каждой используемой метрики по отдельности.

Метрика	SROCC	PLCC	Kendall
PSNR	0,553	0,522	0,366
SSIM	0,525	0,534	0,361
MS-SSIM	0,411	0,451	0,273
FSIM	0,640	0,605	0,450
VSI	0,550	0,574	0,378
GMSD	-0,641	-0,605	-0,450
VIF	0,613	0,596	0,433
LPIPS	-0,390	-0,412	-0,250
CIEDE2000	-0,559	-0,556	-0,391
L2	-0,411	-0,419	-0,274

В рамках дополнительного эксперимента были рассчитаны коэффициенты корреляции между субъективными оценками СОК и каждой из рассматриваемых метрик по отдельности. Результаты представлены в таблице 5.

Анализ полученных данных свидетельствует о том, что наибольшую корреляцию с субъективными оценками демонстрируют метрики FSIM, GMSD и VIF. Результаты подтверждают, что использование ансамблей метрик в сочетании с моделями машинного обучения – такими

как метод опорных векторов (SVM) и полносвязные нейронные сети – способствует значительному повышению точности оценки визуального качества изображений, приближая её к субъективному восприятию наблюдателя. Особенно выраженное улучшение наблюдается при включении формул цветового различия, что подчёркивает их важную роль в задачах объективной оценки визуального качества.

9. ВЫВОДЫ

В данном исследовании рассмотрены подходы к объединению метрик качества изображений с использованием моделей машинного обучения, направленные на оценку визуального качества зашумлённых изображений. Проведено сравнение предсказательной способности различных моделей, а также предложен метод повышения согласованности между модельными оценками и субъективным восприятием путём включения формул цветовых различий. Эффективность данного подхода подтверждается результатами серии экспериментов: наблюдается прирост точности предсказаний в диапазоне от 3% до 18%.

Дополнительно обосновано применение гиперспектрального набора данных для корректного преобразования изображений в цветовые пространства CIELAB и CIELUV на основе колориметрической модели восприятия. Полученные результаты демонстрируют, что интеграция нескольких метрик существенно превосходит по эффективности использование отдельных метрик при оценке качества зашумлённых изображений. Например, прирост корреляции достигает 30% при сравнении метрики FSIM с комбинированным набором метрик (PSNR, FSIM, GMSD, CIEDE2000, L2) для модели SVM и до 112% при сравнении LPIPS с расширенным набором (VIF, GMSD, LPIPS, CIEDE2000, L2) в аналогичной конфигурации.

Использование полного набора метрик, задействованного в данной работе, позволило достичь максимальных значений корреляционных коэффициентов для метода опорных векторов: SROCC = 0,913, PLCC = 0,915, Kendall's τ = 0,736. При этом улучшение по сравнению с лучшей комбинацией метрик для SVM (PSNR, FSIM, GMSD, CIEDE2000, L2) составило порядка 10%, а по сравнению с наилучшей отдельной метрикой (GMSD) – 42,43%.

СПИСОК ЛИТЕРАТУРЫ

1. Wang Z.; Bovik A.; Sheikh H.; Simoncelli E. Image quality assessment: From error visibility to Structural Similarity // *IEEE Trans. Image Process.* 2004. 13(4). pp. 600-612.
2. Zhang L.; Shen Y.; Li H. VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment. // *IEEE Trans. Image Process.* 2014. 23(10). pp. 4270-4281.
3. Ziaei Nafchi H.; Shahkolaei A.; Hedjam R.; Cheriet M. Mean Deviation Similarity Index: Efficient and Reliable Full-Reference Image Quality Evaluator. // *IEEE Access.* 2016. Vol. 4. pp. 5579-5590.
4. Reenu, M., David, D., Raj, S.A., Nair, M. S. Wavelet based sharp features (WASH): an image quality assessment metric based on HVS. // In 2013 2nd International Conference on Advanced Computing, Networking and Security, IEEE. 2013. pp. 79-83.
5. W. Xue, L. Zhang, X. Mou and A. C. Bovik, Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index in *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684-695, Feb. 2014, doi: 10.1109/TIP.2013.2293423.
6. Zhang L.; Mou X.; Zhang D. FSIM: A Feature Similarity Index for Image Quality Assessment // *IEEE Trans. Image Process.* 2011. 20(8). 2378-2386.
7. Zhang X., Feng X., Wang W., Xue W. Edge strength similarity for image quality assessment // *IEEE Signal Process Lett.* 20, pp. 319-322. 2013.

8. Aja-Fernández S, Estépar RS, Alberola-López C, Westin CF. Image quality assessment based on local variance. *Conf Proc IEEE Eng Med Biol Soc.* 2006;2006:4815-8. doi: 10.1109/IEMBS.2006.259516. PMID: 17946653.
9. Wang, Z., Bovik, A.C. A universal image quality index. *IEEE signal processing letters.* 2002. 9(3). pp. 81-84.
10. Ieremeiev, Oleg ; Lukin, Vladimir ; Okarma, Krzysztof et al. / Full-reference quality metric based on neural network to assess the visual quality of remote sensing images. In: *Remote Sensing.* 2020 ; Vol. 12, No. 15.
11. Egiazarian K., Astola J., Ponomarenko N., Lukin V., Battisti F., Carli M. New full-reference quality metrics based on HVS, *CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics, Scottsdale, USA, 2006,* 4 p.
12. N. Ponomarenko, F. Silvestri, K.Egiazarian, M. Carli, V. Lukin, On Between-Coefficient Contrast Masking of DCT Basis Functions, *CD-ROM proceedings of Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07, January, 2007,* 4p.
13. N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, M. Carli, Modified Image Visual Quality Metrics for Contrast Change and Mean Shift Accounting, *Proceedings of CADSM, February 2011, Ukraine,* pp. 305 - 311.
14. S. Bosse, D. Maniry, K. -R. Müller, T. Wiegand and W. Samek, Neural network-based full-reference image quality assessment, 2016 *Picture Coding Symposium (PCS), Nuremberg, Germany, 2016,* pp. 1-5, doi: 10.1109/PCS.2016.7906376.
15. Y. Yuan, Z. Guoqiang, C. Zhenwei and G. Yudong, Color Image Quality Assessment with Multi Deep Convolutional Networks, 2019 *IEEE 4th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 2019,* pp. 934-941, doi: 10.1109/SIPROCESS.2019.8868333.
16. S. Bosse, D. Maniry, K. -R. Müller, T. Wiegand and W. Samek, Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment in *IEEE Transactions on Image Processing,* vol. 27, no. 1, pp. 206-219, Jan. 2018, doi: 10.1109/TIP.2017.2760518.
17. Anustup Choudhury, Scott Daly, Combining Quality Metrics using Machine Learning for improved and robust HDR Image Quality Assessment in *Proc. IS&T Int'l. Symp. on Electronic Imaging: Image Quality and System Performance XVI, 2019,* pp 307-1 - 307-7, <https://doi.org/10.2352/ISSN.2470-1173.2019.10.IQSP-307>.
18. P.P., Konovalenko I.A.; Smagina A.A.; Nikolaev D.P.; Nikolaev. ProLab: A Perceptually Uniform Projective Color Coordinate System. / Konovalenko I.A.; Smagina A.A.; Nikolaev D.P.; Nikolaev P.P. // *IEEE Access.* — 2021.
19. Edul, Sharma Gaurav; Wu Wencheng; Dalal. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. / Sharma Gaurav; Wu Wencheng; Dalal Edul // *Color Research & Application.* — 2005.
20. H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, LIVE image quality assessment database, Release 2, 2005, <http://live.ece.utexas.edu/research/quality>.
21. E. C. Larson and D. M. Chandler. (2009). Consumer Subjective Image Quality Database. Available: <http://vision.okstate.edu/index.php>.
22. N. Ponomarenko et al., Color image database TID2013: Peculiarities and preliminary results in *Proc. 4th Eur. Workshop Vis. Inf. Process. (EUVIP), 2013,* pp. 106–111.
23. N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics, *Advances of Modern Radioelectronics,* Vol. 10, pp. 30-45, 2009.
24. Wang, Z. Multiscale structural similarity for image quality assessment / Z. Wang, Eero Simoncelli, Alan Bovik // *Conference Record of the Asilomar Conference on Signals, Systems and Computers.* — 2003.

25. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," in *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430-444, Feb. 2006, doi: 10.1109/TIP.2005.859378. keywords: Distortion measurement;Image quality;Humans;Quality assessment;Algorithm design and analysis;Predictive models;Psychology;Visual system;Video sharing;Statistics;Image information;image quality assessment (QA);information fidelity;natural scene statistics (NSS),
26. Zhang, Richard and Isola, Phillip and Efros, Alexei A and Shechtman, Eli and Wang, Oliver, The Unreasonable Effectiveness of Deep Features as a Perceptual Metric // *CVPR* - 2018.
27. Commission Internationale de l'Éclairage (CIE), CIE Publication No. 15: Colorimetry // 1976.
28. Mohd. Ameen. (2016). An Extensive Review of Medical Image Denoising Techniques. *Global Journal of Medical Research*, 16(D2), 23–27.
29. Oguzhanoglu, S., Kapucuoglu, I., and Sunar, F.: COMPARISON OF SATELLITE IMAGE DENOISING TECHNIQUES IN SPATIAL AND FREQUENCY DOMAINS, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B3-2022, 1241–1247, <https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-1241-2022>, 2022.
30. Okarma, K.; Lech, P.; Lukin, V.V. Combined Full-Reference Image Quality Metrics for Objective Assessment of Multiply Distorted Images. *Electronics* 2021, 10, 2256. <https://doi.org/10.3390/electronics10182256>
31. Chiu, C.-C.; Lin, C.-Y.; Chi, Y.-J.; Hsu, H.-H.; Chen, P.-H.; Jiang, H. A Convolutional Neural Network with Multifrequency and Structural Similarity Loss Functions for Electromagnetic Imaging. *Sensors* 2024, 24, 4994. <https://doi.org/10.3390/s24154994>
32. B. Arad, Z. Ben-Haim, R. Margulis, Y. Atzmon, R. Giryas, and R. Timofte, "NTIRE 2022 Spectral Recovery Challenge and Data Set," *Proc. CVPRW*, pp. 436–445, 2022.
33. Ershov, E., Panshin, A., Ermakov, I., Banić, N., Savchik, A., Bianco, S. (2024). Reliability and Stability of Mean Opinion Score for Image Aesthetic Quality Assessment Obtained Through Crowdsourcing. In *Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (pp.365-372). Science and Technology Publications, Lda [10.5220/0012462000003660].
34. CIE (1931). *Commission Internationale de l'Éclairage Proceedings*, Vol. 3, Cambridge University Press.

Creating a neural network noise reduction metric based on human color perception

D. E. Abramov, S. A. Korchagin, E. E. Zaichenkova, E. I. Ershov

The presence of noise in images is one of the most significant challenges in modern computational photography, which explains the rapid and continuous development of this field. One of the most complex issues is finding the optimal image quality assessment (IQA) method. Currently, both analytical IQA metrics, such as PSNR and SSIM, and neural network-based metrics, such as LPIPS and CLIP-IQA, are in use. Interestingly, no existing IQA methods take into account knowledge of the color perception characteristics of the standard CIE XYZ observer. This work is the first to demonstrate that incorporating such knowledge using the CIEDE2000 color difference formula and Euclidean distance in the CIELUV space can improve the accuracy of image quality assessment. We propose a new IQA metric based on a combination of classical and color-based formulas for comparing images. Experiments on a new dataset show an increase in correlation with human opinion scores by up to 20%, confirming the high potential for further development in this direction.

KEYWORDS: denoising metrics, human color perception, Mean Opinion Score (MOS), color spaces.