

Новый датасет и выявление ключевых признаков, определяющих качество стилизации изображений на его основе

Китов В. В.^{1,2*}, Абрамов В. А.¹, Ахтырченко М. К.¹

^{1*} ф-т ВМК, МГУ имени М.В.Ломоносова, Ленинские горы, д.1, стр. 52, 119991, Москва, Россия.

² дирекция по науке и инновациям, РЭУ имени Г.В.Плеханова, Стремянный пер., д. 36, Москва, 109992, Москва, Россия.

* Автор(ы), ответственный(ые) за переписку. E-mail(s):

v.v.kitov@yandex.ru;

Соавторы: ab.val26@yandex.ru; ahm.misha@gmail.com;

Аннотация

В работе предлагается новый набор данных, предназначенный для совершенствования нейросетевой стилизации изображений, представляющей собой автоматическую перерисовку одного изображения в стиле другого. Набор данных охватывает изображения с различным контентом и стилем разного размера и содержит 10000 стилизованных изображений, вручную оцененных тремя аннотаторами по шкале от 1 до 10. На основе полученных оценок в работе определяется, какие факторы в наибольшей степени отвечают за положительное и отрицательное восприятие стилизации пользователями, и выявляются количественные показатели, оказывающие статистически значимое влияние на их оценки. Приводится методология создания предложенного набора данных, который может быть использован для автоматизации различных задач, связанных с настройкой стилизующих алгоритмов и оценкой их качества.

Ключевые слова: Нейросетевой перенос стиля, датасет, компьютерное зрение, нейронные сети, качество стилизации.

Получено редакцией 11.11.2025; внесены авторские правки 20.11.2025; принята к публикации 21.11.2025

1 Введение

Перенос стиля — это увлекательная область исследований, в которой для заданного изображения *контента* (например, семейной фотографии) и изображения *стиля* (например, картины известного художника) решается задача автоматической перерисовки контента в заданном стиле, получив в результате целевую *стилизацию*, как показано на рис. 1.

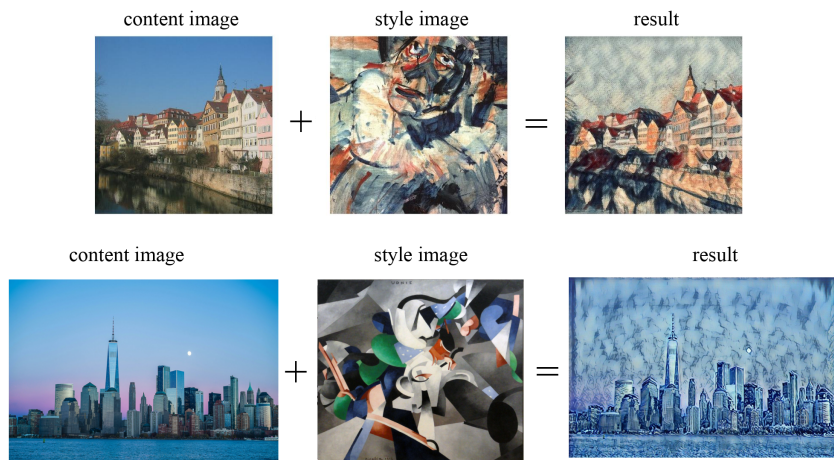


Рис. 1: Иллюстрация стилизации изображения контента по изображению стиля с переносом цветов стиля (верхний ряд) и с сохранением цветов исходного контентного изображения (нижний ряд).

Перенос стиля автоматизирует создание произведений цифрового искусства, и может применяться в дизайне, моде и рекламе. Он также способствует коммуникации, позволяя делиться яркими и запоминающимися изображениями в социальных сетях. Многие мобильные приложения и онлайн-платформы предлагают услуги переноса стиля, такие как prisma-ai.com, ostagram.me и picsart.com.

Но что делает стилизацию хорошей или плохой в восприятии пользователя? Поскольку конечный результат — это цифровое произведение искусства, сложно дать формальный показатель качества, и большинство исследований полагаются на человеческие оценки при сравнении различных методов стилизации. В таких оценках пользователей-ассессоров просят оценить стилизации по абсолютной шкале или выбрать лучшую из нескольких вариантов. Программа оценки качества стилизаций при этом разрабатывается заново каждой исследовательской группой.

Набор изображений контента и стиля индивидуален для каждого исследования, что делает проведенные эксперименты невозпроизводимыми и несопоставимыми между статьями.

Для решения этих проблем мы предлагаем большой набор данных для переноса стиля¹ с разрешительной лицензией, содержащий контент, стили разных размеров и полученные на их основе стилизации, качество которых было вручную оценено тремя ассессорами. В рамках разметки представленного датасета была разработана программа для оценки качества стилизаций, код которой публично доступен².

На представленном датасете можно обучать нейросетевые методы, которые бы оценивали качество стилизации автоматически, упрощая сравнение стилизующих методов друг с другом.

Представленный набор данных также можно использовать для обучения моделей, решающих следующие задачи:

- предсказать, насколько хорошо заданные контент, стиль и его размер соответствуют друг другу;
- предсказать оптимальный размер стиля по заданному контенту и стилю;
- по заданному контенту и последовательности стилей ранжировать стили вместе с рекомендуемыми размерами от наиболее подходящих и соответствующих друг другу к наименее подходящим.

Соответствие при этом определяется как ожидаемая оценка качества результирующей стилизации по известным входным данным.

Такие задачи полезны в практических приложениях переноса стилей, так как они автоматизируют выбор наилучшего изображения стиля и его размера для пользовательского контента.

Изображения контента, составляющие набор данных, были тщательно отобраны, чтобы охватить широкий набор изображений, содержащих разнообразные объекты, показанные с разных позиций и при разном освещении. Стили были взяты из репозитория GitHub, содержащих основные методы переноса стилей. Стилизации оценивались по абсолютной шкале от 1 (низшее качество) до 10 (высшее качество). Поскольку оценка художественных работ субъективна, каждое изображение оценивалось тремя аннотаторами. Наш набор данных содержит в общей сложности 30 000 оценок.

Распространённые методы переноса стилей отображают не только художественные узоры, такие как мазки кисти, но и распределение цвета на изображении стиля. Примеры переноса стиля с учётом и без учёта цветовой информации показаны на рис. 1.

При переносе стиля мы концентрируемся только на переносе художественных паттернов без переноса цветов, поскольку перекрашивание

¹<https://github.com/victorkitov/style-transfer-dataset>

²<https://github.com/EnriFermi/image-evaluation-app>

цветов может быть достигнута отдельными хорошо зарекомендовавшими себя методами, такими как сопоставление среднего значения и дисперсии цветовых каналов (mean and variance matching) или сопоставление гистограмм (histogram matching) [1]. Поэтому мы оценили чистый перенос стиля, отображающий только художественные узоры без цветовой информации, чтобы исключить смещение оценок, вызванное предпочтением определённых цветов пользователями.

Также важным фактором является размер изображения стиля, поскольку все методы стилизации основаны на свёртках с фиксированными ядрами. Это упоминалось в [2] и показано на рис. 2. Увеличение размера стилевого изображения накладывает более крупные художественные паттерны на изображение контента, поэтому при создании датасета были рассмотрены различные размеры стилей.

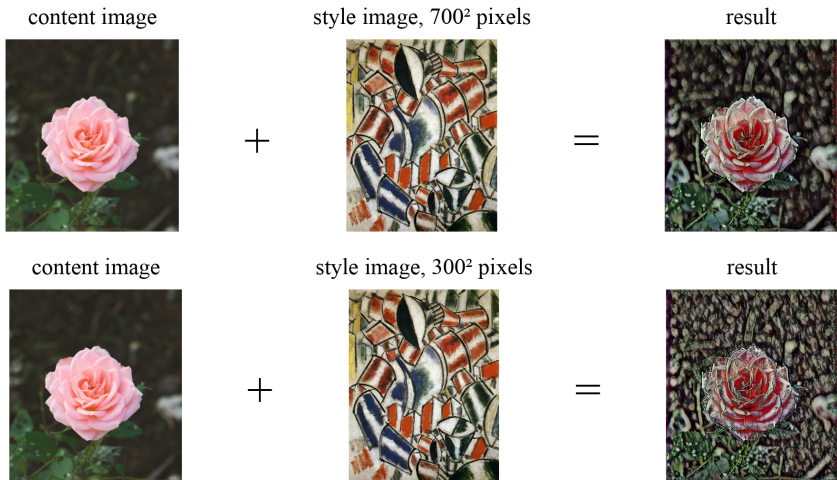


Рис. 2: Перенос стиля с использованием стилевого изображения разного размера, состоящего из 700^2 пикселей (верхний ряд) и 300^2 пикселей (нижний ряд).

Стилизации выполнялись с помощью популярного метода ArtFlow [3], обеспечивающего схожие на качественном уровне результаты с другими *современными* методами, как показано в попарных сравнительных визуализациях [4]. Сходство методов передачи стиля объясняется тем, что, несмотря на использование разных архитектур, их параметры обучаются на одной и той же функции потерь, предложенной в [5], или на её небольших вариациях, таких как сближение централизованных ковариационных матриц между признаками во внутренних представлениях вместо нецентрированных или сближение только средних значений и

стандартных отклонений вместо полных ковариационных матриц. Все эти модификации решают на качественном уровне одну и ту же задачу: приближение распределения признаков стилизованного изображения к распределению признаков изображения стиля, поэтому приводит к похожим результатам.

Применимость нашего набора данных несколько ограничена для *более ранних* методов стилизации, основанных на других принципах и использующих другие функции потерь, а именно:

- перенос стиля на основе оптимизации [5], где оптимизация выполняется не в пространстве весов нейронной сети, а в пространстве пикселей результирующей стилизации;
- перенос стиля на основе патчей [6],[7], где стилизация выполняется путём замены фрагментов контента наиболее похожими фрагментами стиля.

Мы делимся нашими подходами для создания разнообразного и информативного датасета для переноса стиля и призываем научное сообщество расширять существующий.

Учитывая разнообразие оцененных стилизаций, мы делимся своими наблюдениями об эффективности и ограничениях переноса стиля для разных входных данных, а также описываем требования на данные, способствующие генерации качественных стилизаций.

В работе анализируются распространённые артефакты стилизации и выявляются особенности изображений, оказывающие статистически значимое влияние на восприятие и оценку качества стилизаций пользователями. Мы надеемся, что представленные результаты укажут направления для дальнейшего совершенствования методов переноса стиля.

В целом, вклад нашей работы заключается в следующем:

- представлен свободно доступный датасет, который может быть использован для понимания того, как работает перенос стиля в различных контекстах, и может быть использован для построения моделей, автоматически оценивающих качество результирующей стилизации, а также ожидаемое качество ещё не сгенерированной стилизации для заданного контента, стиля и его размера;
- выложена в открытый доступ программа для разметки качества стилизаций;
- описана методология создания разнообразного и информативного датасета для переноса стиля;
- мы делимся общими наблюдениями о том, при каких условиях методы переноса стиля работают хорошо, а при каких — плохо.
- найдены количественные характеристики изображений, оказывающие статистически значимое влияние на итоговую пользовательскую оценку качества стилизаций.

2 Обзор литературы

Насколько нам известно, была предпринята только одна попытка создания большого общедоступного набора данных для задач стилизации в [8]. Хотя мы признаём важность этой работы, на наш взгляд, она имеет ограничения, которые мы попытались преодолеть в нашем исследовании:

- Было оценено всего 150 пар изображений контента и стиля, чего недостаточно для обучения даже простых нейронных сетей. Наш набор данных содержит оценки 10 000 пар контента и стиля.
- Изображения контента были масштабированы и обрезаны до разрешения 512×512 . Мы утверждаем, что конечный пользователь заинтересован в стилизации изображений контента более высокого разрешения без ограничения, согласно которому изображение должно быть квадратным. Таким образом, изображения контента в нашем наборе данных имеют высокое разрешение и сохраняют исходное (произвольное) соотношение сторон.
- Размер изображений стиля также был изменен до 512×512 . Хотя это распространённая стратегия масштабирования изображений стилей до такой формы в научных статьях, конечный пользователь заинтересован в применении методов стилизации к стилям произвольного размера и формы. Кроме того, мы подчеркиваем, что размер стиля оказывает существенное влияние на качество итоговой стилизации, как показано на рис.2. Поэтому мы не накладываем никаких ограничений на соотношение сторон и рассматриваем различные размеры каждого изображения стиля, содержащие 150^2 , 300^2 , 500^2 и 700^2 пикселей. Мы показываем, что размер изображения стиля влияет на размер наложенных шаблонов стиля и оказывает существенное влияние на качество получаемого результата.
- Пары «контент-стиль» в [8] были тщательно подобраны для соответствия друг другу (семантически на них должны быть изображены похожие объекты). Однако реальные пользователи могут сочетать произвольные изображения контента и стиля, и часто это приводит к положительным результатам, даже если исходное изображение контента и стиля не соответствуют друг другу на семантическом уровне, как можно видеть на примере многих стилизаций сайта ostagram.me.
- Методы переноса стиля, оцениваемые в [8], были предложены в 2017–2020 годах. Мы рассматриваем только метод ArtFlow [3], предложенный в 2021 году, который даёт качественно схожие результаты с более поздними методами, что можно увидеть в визуальных сравнениях [4], а также подтверждается нашими собственными экспериментами.
- В [8] аннотаторам было предложено выбрать лучшую стилизацию из двух альтернатив. Это даёт лишь N результатов при N парных сравнениях. В нашем же исследовании мы попросили каждого аннотатора

оценить каждую стилизацию индивидуально по шкале от 1 до 10, что позволяет из N индивидуальных оценок получить $N(N-1)/2$ парных сравнений, сравнивая каждое изображение с каждым по абсолютной шкале.

3 Процесс создания датасета

Далее будет описан процесс создания предлагаемого набора данных.

3.1 Изображения контента и стиля

Набор данных был создан с использованием 50 изображений контента и 50 изображений стиля. Поскольку размер стиля сильно влияет на результат стилизации, как показано на рис. 2, изображения стилей были масштабированы (с сохранением соотношения сторон) до 150^2 , 300^2 , 500^2 и 700^2 пикселей. Мы не использовали предварительное сопоставление изображений контента и стиля, поскольку часто даже семантически разные изображения контента и стилей дают выразительные результаты. Вместо этого мы стилизовали каждую комбинацию контента, стиля и размера стиля, получив $50 \cdot 50 \cdot 4 = 10000$ стилизаций. В конечном итоге некоторые комбинации хорошо сочетаются друг с другом, некоторые — нет, и наш набор данных позволяет оценить (и обучить модели автоматической оценки), насколько хорошо контент и стиль соответствуют друг другу, и какой размер стиля является оптимальным.

Поскольку художественные вкусы разных людей различаются, каждая стилизация была оценена тремя аннотаторами, что в итоге дало 30000 оценок. Это позволяет оценить уровень разногласий между разными аннотаторами.

Изображения контента включают в себя разнообразные фотографии с людьми, животными, зданиями и другими распространёнными объектами, такими как поезда, автомобили, самолёты, деревья и цветы. Изображения были выбраны с учётом различных условий освещения (день, вечер, ночь) и различных расстояний до объектов — близкого, среднего и дальнего. Также были включены пейзажи и виды улиц.

Большинство изображений контента были загружены с сайта unsplash.com по разрешительной лицензии³. Поскольку неудачный перенос стиля ухудшает качество фотографий, портретные фотографии людей были созданы с помощью генеративно-состязательной сети [9] (контенты № 14 и 17) и диффузионной модели [10] (контенты № 8, 13, 22, 26, 48) и масштабированы⁴ без потери фотореализма. Все изображения контента были масштабированы с сохранением соотношения сторон до 900^2 пикселей, исходя из того, что окончательное применение переноса стиля выполняется в HD-разрешении.

³<https://unsplash.com/license>

⁴Масштабирование с помощью <https://www.pixelcut.ai/image-upscaler>

Изображения стилей были взяты из репозитория Github, содержащих популярные методы переноса стилей, в основном ArtFlow⁵, поскольку эти изображения широко известны и распространены в сообществе, разрабатывающем методы стилизации, и могут свободно использоваться в исследовательских целях.

3.2 Создание стилизации

После получения 50 изображений контента и 50 изображений стилей в 4 различных масштабах было создано 10 000 стилизованных изображений с помощью алгоритма ArtFlow [3] с использованием официальной предобученной модели⁶. Эта модель обеспечивает высокое качество результатов, работает с контентом и стилями разных размеров с произвольным соотношением сторон и достаточно экономична по вычислительным ресурсам (способна выдавать результаты высокого разрешения на видеокарте с 11 ГБ памяти).

Реализации более современных алгоритмов переноса стилей, а именно AdaAttN [11]⁷ и Stytr2 [4]⁸, также способны выдавать результаты высокого качества, однако

1. более требовательны к вычислительным ресурсам (11 ГБ памяти видеокарты уже недостаточно для получения результатов в высоком разрешении),
2. официальные реализации работают только при одинаковом размере контента и стиля (пользователю же удобнее работать с изображениями произвольных размеров),
3. требуют, чтобы изображение контента и стиля имело квадратное соотношение сторон (фактические же соотношения сторон могут быть прямоугольными).

Из-за этих ограничений мы придерживаемся метода ArtFlow в нашем анализе. Кроме того, сравнительный визуальный анализ стилизаций, полученных с помощью ArtFlow, AdaAttN и Stytr2, представленный в [4], показывает схожие результаты на качественном уровне.

Мы не рассматривали генеративно-сопоставительные сети и модели диффузии, способные переносить стиль, поскольку они ещё более требовательны к вычислительным ресурсам, а нас интересовал анализ и повышение качества более легких классических методов стилизации изображений.

⁵<https://github.com/pkuanjie/ArtFlow/tree/main/data/style>

⁶<https://github.com/pkuanjie/ArtFlow>

⁷<https://github.com/Huage001/AdaAttN>

⁸<https://github.com/diyiyiii/StyTR-2>

3.3 Перекрашивание стиля

По умолчанию методы стилизации переносят не только стилевые паттерны (такие как мазки кисти, эффекты размытия и т.д.), но и распределение стиливых цветов. Это может привести к нестабильным результатам, например, к перерисовке яркой фотографии в тёмные цвета. Это также может привести к цветовой предвзятости в оценках пользователей (некоторые цвета более предпочтительны для респондентов). Чтобы избежать обеих проблем, мы осуществляем стилизацию с сохранением цветов исходного контентного изображения. Для этого изображение стиля перекрашивается в цвета контента до стилизации, используя сопоставление среднего и дисперсии в цветовом пространстве LAB, а именно с использованием следующего алгоритма:

1. Преобразовать изображение контента и стиля из цветовой схемы RGB в LAB.
2. Рассчитать значения $\mu_l^c, \mu_a^c, \mu_b^c / \mu_l^s, \mu_a^s, \mu_b^s$ для цветовых каналов LAB изображения контента/стиля.
3. Рассчитать стандартные отклонения $\sigma_l^c, \sigma_a^c, \sigma_b^c / \sigma_l^s, \sigma_a^s, \sigma_b^s$ для цветовых каналов LAB изображения контента/стиля.
4. Изменить масштаб цветовых каналов LAB l^s, a^s, b^s , чтобы приблизить распределение цвета стиля к цветам контента:

$$l^s := \frac{\sigma_l^c}{\sigma_l^s}(l^s - \mu_l^s) + \mu_l^c$$

$$a^s := \frac{\sigma_a^c}{\sigma_a^s}(a^s - \mu_a^s) + \mu_a^c$$

$$b^s := \frac{\sigma_b^c}{\sigma_b^s}(b^s - \mu_b^s) + \mu_b^c$$

Перекрашивание выполнялось в цветовом пространстве LAB, поскольку это цветовое пространство более естественно: равномерные изменения компонентов цвета LAB соответствуют равномерным изменениям воспринимаемого человеческим глазом цвета⁹. Мы также экспериментировали с перекрашиванием методом сопоставления гистограмм¹⁰, используя библиотеку `skimage`, но этот метод показал результаты худшего качества и далее не рассматривался.

⁹https://en.wikipedia.org/wiki/CIELAB_color_space

¹⁰https://en.wikipedia.org/wiki/Histogram_matching

3.4 Процесс разметки

Каждая из 10 000 стилизаций была размечена тремя аннотаторами по шкале от 1 (худший) до 10 (лучший), что дало 30 000 оценок с помощью программы, специально разработанной для оценки изображений человеком.

В процессе разметки стилизации демонстрировались каждому аннотатору на весь экран. Аннотаторам предлагалось охватить весь диапазон оценок от 1 до 10 для всего набора данных, однако ограничений по охвату этого диапазона для конкретного контента и стиля не было. Действительно, некоторые контенты и стили в среднем давали хорошие результаты, а некоторые — плохие.

Для оценки были представлены только стилизации, поскольку конечный пользователь методов стилизации оценивает лишь конечный результат. Аннотаторам было позволено оценивать изображения в соответствии с их личными эстетическими вкусами, «готовностью использовать их в качестве картины на стене» или «иллюстрации на веб-сайте определённой тематики». Аннотаторам предлагалось не учитывать сам контент, а учитывать только художественную выразительность изображения и его привлекательность. Для определённых стилей стилизация давала фотореалистичный результат без какого-либо художественного эффекта. В таких случаях аннотаторам предлагалось оценить стилизацию в диапазоне от 4 до 6 (поскольку заметная стилизация не состоялась), в зависимости от насыщенности цвета и наличия локальных артефактов переноса стиля.

Каждый аннотатор сначала оценивал все стилизации одного изображения контента, затем другого и т.д. Это было сделано для того, чтобы аннотаторам было проще сравнивать относительное качество стилизации одного и того же изображения контента. После каждой оценки поверх оцениваемого изображения на 0.5 секунды отображалась большая оценка, чтобы помочь аннотаторам запомнить, как именно они оценили каждый вид стилизации. Порядок следования изображений контента для разных аннотаторов был случайным для компенсации гипотетических тенденций в оценках по мере продвижения процесса при оценивании.

Было замечено, что ArtFlow часто создает артефакты по краям стилизованного изображения, как показано на рис. 3.

Аннотаторам было предложено не снижать оценку стилизации из-за наличия этого конкретного артефакта, поскольку его можно легко удалить, слегка обрезав края изображений. За все остальные артефакты оценка снижалась.



Рис. 3: Распространённый артефакт вдоль границ стилизаций, выполненных стилизующим методом ArtFlow.

4 Общий анализ оценок

Распределение индивидуальных оценок каждого из трёх аннотаторов, а также распределение средних оценок (для каждой стилизации) показаны на рис. 4. Средняя оценка имеет асимметрию в сторону более высоких оценок. Индивидуальные оценки распределены по-разному, что подтверждает гипотезу о небольших различиях в эстетических предпочтениях разных людей. Стандартное отклонение оценок в основном не превышает двух баллов, как показано на рис. 5.

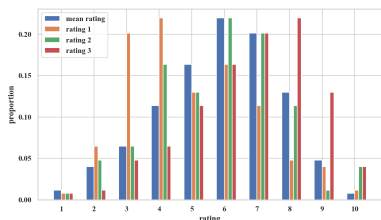


Рис. 4: Распределение оценок для 3 аннотаторов и средняя оценка.

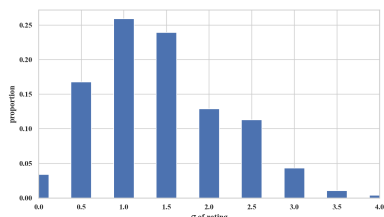


Рис. 5: Распределение стандартных отклонений оценок для каждой стилизации по аннотаторам.

Для анализа корреляций между оценками использовалась ранговая корреляция Кендалла¹¹, способная оценивать силу монотонных зависимостей между оценками независимо от типа этих зависимостей. Мы использовали реализацию меры Тау-В Кендалла из библиотеки `scipy`¹². Для проверки статистической значимости мы провели статистический тест с нулевой гипотезой о том, что ранговая корреляция равна нулю с двусторонней альтернативой [12] [13], которая была отвергнута с высокой степенью достоверности ($\ll 0.01$). Корреляции в таблице 1 показывают, что оценки каждого аннотатора имеют в значительной степени монотонную зависимость друг от друга, и, таким образом, мы

¹¹https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient

¹²<https://scipy.org/>

можем использовать усредненные оценки для оценки общего качества каждой стилизации.

Таблица 1: Корреляции Кендалла (Тан-В) между рейтингами трёх аннотаторов.

	аннотатор 1	аннотатор 2	аннотатор 3
аннотатор 1	1.0000	0.4030	0.3612
аннотатор 2		1.0000	0.4314
аннотатор 3			1.0000

В таблицах 5 и 6 показано, что качество стилизации сильно зависит от изображения контента: некоторые изображения контента имеют в среднем более низкий средний рейтинг, чем другие, поскольку хуже поддаются стилизации.

5 Факторы влияния

В этом разделе анализируются наиболее важные, на наш взгляд, факторы, определяющие оценки пользователей.

5.1 Цветовое разнообразие и резкость

Высоко оцененные стилизации имеют несколько общих черт: они слабо искажают формы объектов на изображении, но добавляют локальные текстуры из изображения стиля к фону. Качество стилизации также зависит от цвета и разнообразия в яркости полученного изображения. Для вычисления разнообразия мы преобразуем стилизацию в цветковое пространство LAB и вычисляем яркостное разнообразие как стандартное отклонение канала яркости $\sigma(L)$, а цветовое разнообразие — как $\sqrt{\sigma(L)^2 + \sigma(A)^2 + \sigma(B)^2}$, учитывая все три канала. Коэффициенты корреляции Тан-В Кендалла в таблице 2 показывают, что корреляция между средними оценками и этими показателями существует и является статистически значимой (отвергнутая нулевая гипотеза заключалась в отсутствии какой-либо зависимости).

Таблица 2: Корреляции Тан-В Кендалла между средним рейтингом и различными показателями. Все соответствующие р-значения значительно меньше 0.01 (т.е. результаты статистически значимы).

мера	$\sigma(L)$	$\sqrt{\sigma(L)^2 + \sigma(A)^2 + \sigma(B)^2}$	средняя резкость
Тан-В Кендалла	0.0846	0.0941	-0.2258
р-значение	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$

Средняя резкость стилизации также влияет на оценку её качества. Мы вычисляем резкость изображения как дисперсию лапласиана

изображения (результата применения дискретного фильтра Лапласа к изображению), как предложено в работе [14]. Корреляция этой меры с оценками отрицательная, как показано в таблице 2. Это означает, что для слишком резких стилизаций средняя оценка их качества снижается.

Излишне резкие изображения обычно получались с использованием стилизованных изображений, содержащих чрезмерное количество мелких и резких узоров. Эти узоры перегружали результат стилистическими деталями, затрудняя восприятие исходного контента, как показано на рис. 10 (средняя строка).

5.2 Размытие текстуры и воспроизведение границ

Стилизации с низкими оценками показывают, что процесс переноса стиля сильно зависит от изображений контента. На рисунке представлены изображения с самыми низкими средними оценками. Портреты в анфас оцениваются ниже, чем другие изображения. Это происходит из-за искажения черт лица, вызванного переносом стиля, особенно на коже человека — даже минимальные морщины усиливаются переносом стиля, как видно на двух верхних рядах рис. 6.

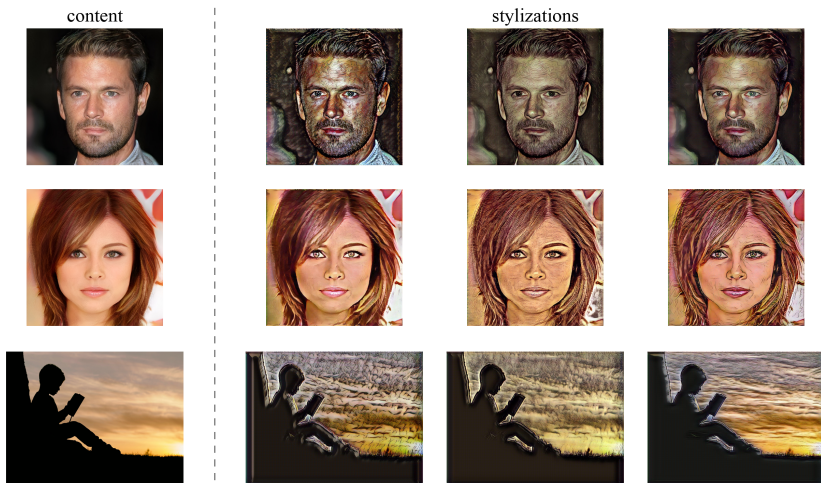


Рис. 6: Примеры изображений контента, дающие стилизации с низкими оценками. Примеры стилизаций на их основе.

Ещё одной распространённой особенностью плохо стилизованного контента является наличие равномерно или плавно окрашенных областей, таких как чёрная фигура и небо в нижнем ряду рис. 6, которые

часто искажаются случайными стилистическими узорами, что делает их негладкими. Хорошо поддаются стилизации контенты, лишённые отмеченных особенностей, такие как на рис. 7.



Рис. 7: Примеры изображений контента, обеспечивающие стилизацию с высокими оценками. Примеры стилизаций на их основе.

5.3 Воспроизведение лиц

Изображения человеческих лиц, как правило, очень чувствительны к переносу стиля. Минимальные морщины имеют тенденцию усиливаться чёткими и контрастными стилями, а небольшие изменения черт лица приводят к снижению его узнаваемости. Мы обнаружили сильную корреляцию Кендалла, равную 0.4696, между оценками качества и средней косинусной мерой близости между эмбедингами лиц, извлечённых с изображения контента и соответствующей стилизации.

Нулевая гипотеза об отсутствии корреляции имеет пренебрежимо малое р-значение, что подтверждает, что воспроизведение исходного лица необходимо для высокой оценки качества стилизации.

Эмбединги были получены с помощью модели VGG-face, предложенной [15] и реализованной в библиотеке DeepFace¹³.

5.4 Размер и характеристики стиля

Среднее качество стилизации сильно зависит от стиля и его размера, как показано в таблицах 4 и 3, иллюстрирующих средние оценки для

¹³<https://github.com/serengil/deepface>

Таблица 3: 5 стилей с наименьшим рейтингом в наборе данных.

Стиль №	7	7	7	38	7
Размер стиля (px)	150 ²	300 ²	500 ²	700 ²	700 ²
средний рейтинг	1.87	2.02	2.10	2.24	2.28
стандартное отклонение	1.06	1.25	1.28	0.84	1.25

Таблица 4: 5 стилей с наибольшим рейтингом в наборе данных.

Стиль №	43	41	19	43	43
Размер стиля (px)	300 ²	700 ²	300 ²	700 ²	500 ²
средний рейтинг	7.85	7.86	7.92	7.98	8.28
стандартное отклонение	1.28	1.38	1.26	1.27	1.11

пяти стилей, генерирующих стилизации самого лучшего и худшего качества. Эта зависимость гораздо сильнее, чем от изображения контента (таблицы 6 и 5).

Таблица 5: Пять изображений контента с наибольшим рейтингом в наборе данных.

Контент №	4	20	5	3	26
средний рейтинг	6.44	6.46	6.60	6.62	6.65
стандартное отклонение	1.34	1.26	1.42	1.22	1.07

Таблица 6: Пять изображений контента с наименьшим рейтингом в наборе данных.

Контент №	14	17	36	43	19
средний рейтинг	4.60	4.64	4.90	4.91	5.06
стандартное отклонение	1.41	1.50	1.48	2.20	1.51

Сильное влияние размера стиля показано в таблице 7 и на рис. 2. Размер изображения стиля влияет на размер художественных эффектов, передаваемых с изображения стиля. Результаты в таблице 7 показывают, что слишком малый размер негативно сказывается на качестве, как и слишком большой размер. Оптимальный размер стиля составляет 500² пикселей, обеспечивая наилучшие результаты в среднем.

Таблица 7: Размеры стилей и их соответствующие статистики оценок.

размер стиля (пикс.)	150 ²	300 ²	500 ²	700 ²
средний рейтинг	5.19	5.95	6.11	6.09
стандартное отклонение	1.43	1.46	1.51	1.55

Малый размер стиля приводит к небольшим высокочастотным художественным эффектам, которые подавляют восприятие исходного контента, как показано на рис. 2. С другой стороны, если размер стиля слишком большой, свертки, используемые в методах переноса стиля, не способны воспроизводить крупные художественные узоры, поскольку у них фиксированные рецептивные области, и перенос стиля оказывает незначительное влияние на изображение контента, как показано в нижней строке рис. 8.

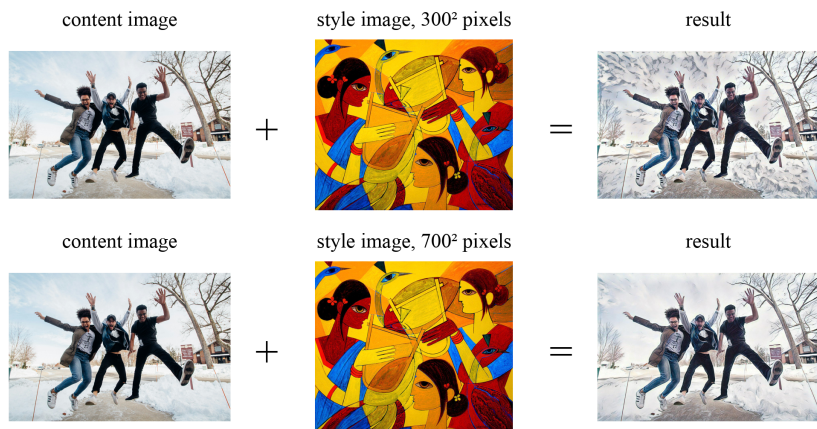


Рис. 8: Перенос стиля не позволяет передать крупные художественные узоры для изображений стиля большего размера.

Комбинации стиля и размера с наивысшими средними оценками представлены на рис. 9. В таких случаях размер изображения стиля не слишком мал (минимальный размер составляет 300^2 пикселей), а само изображение содержит художественные узоры, которые не слишком малы (чтобы не перегружать контент) и не слишком велики (чтобы их могли воспроизвести свертки с фиксированной рецептивной областью).

Изображения в благоприятном стиле также содержат разнообразные цвета и разнообразные художественные узоры под разными углами и разной формы. Разнообразие углов позволяет точно сохранить границы изображения, сохраняя его узнаваемость. Крупные и мелкие формы художественных узоров позволяют хорошо стилизовать высокочастотный контент (на переднем плане) и низкочастотный контент (на фоне).

Стили же с низким средним рейтингом, например, показанные на рис. 10, лишены разнообразия цветов, граней и более сложных текстур (первый и последний стили). Это приводит к чрезмерному сглаживанию на стилизации и размытию деталей контента. Второй неблагоприятный



Рис. 9: Примеры стилей, дающих стилизации с высокими оценками. Размеры стилизованных изображений — 300^2 , 500^2 и 700^2 пикселей соответственно. Примеры стилизаций на их основе.

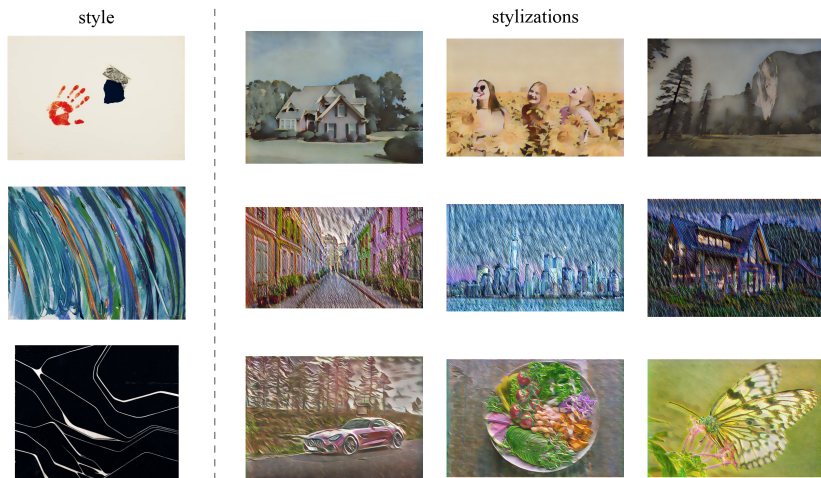


Рис. 10: Примеры стилей, обеспечивающих стилизации с низкими оценками. Размеры стилизованных изображений — 150^2 , 300^2 и 500^2 пикселей соответственно. Примеры стилизаций на их основе.

случай, показанный во второй строке рис. 10, возникает, когда изображение стиля содержит слишком много высокочастотных паттернов

под одним углом. Перенесенные высокочастотные паттерны перегружают исходный контент. Происходит дополнительная потеря информации о контенте, поскольку границы стиля находятся под одним углом, в то время как объекты контента содержат углы, повернутые различным образом. Таким образом, большинство из них игнорируются и стираются при переносе стиля, что ухудшает узнаваемость исходного контента.

6 Рекомендации для успешной стилизации

Приведём итоговый список рекомендаций для создания качественных стилизаций методами нейросетевого переноса стиля.

1. Размер стиливых паттернов сильно влияет на качество стилизации. Если они слишком малы, они становятся неузнаваемыми, и перенос стиля сводится к фотореализму. Если паттерны содержат много мелких чётких контуров, их перенос зашумляет исходную информацию о контенте. Кроме того, алгоритмы переноса стиля имеют ограниченные возможности для воспроизведения крупных художественных узоров, распределенных по обширным областям. Стилиевые паттерны с крупными плавно меняющимися узорами не переносятся, а лишь чрезмерно сглаживают конечную стилизацию.
2. Эффективный стиль должен содержать комбинацию крупных и мелких художественных узоров, и они должны приблизительно соответствовать распределению высокочастотных и низкочастотных данных на изображении контента во избежание их зашумления и сглаживания.
3. Размер самого изображения стиля важен, поскольку он определяет средний масштаб художественных узоров на изображении. Таким образом, алгоритмы переноса стиля должны уметь работать с изображениями стилей разных размеров.
4. Изображение стиля должно содержать разнообразные текстуры и цвета. Использование слишком простых стилей приводит к размытию и потере важных деталей при стилизации.
5. Изображение стиля должно содержать границы под разными углами. В этом случае перенос стиля сохраняет возможность сохранения сложных контуров изображения контента, тем самым сохраняя его узнаваемость.
6. Перенос стиля имеет тенденцию преувеличивать даже минимальные изменения цветов, что приводит к заметным артефактам на больших гладких объектах, таких как небо. Также стилизация усиливает даже минимальные морщины на лицах, что ограничивает применимость метода для стилизации лиц крупным планом.

7 Заключение

В работе представлен новый датасет для задач стилизации изображений, охватывающий различные изображения контента и стилей разного размера, содержащий 10 000 стилизованных изображений, оцененных тремя аннотаторами. Этот набор данных может использоваться для обучения моделей, оценивающих ожидаемое качество стилизации по ней самой либо по входным изображениям контента и стиля. Описана методология формирования датасета, а программа для оценки качества изображений выложена в открытый доступ.

На основе полученных оценок проанализированы количественные и качественные факторы, влияющие на качество получаемых стилизаций, на основе которых сформулированы рекомендации по выборке стилизованных изображений и подбору их размера для получения наилучших результатов.

Мы надеемся, что полученные выводы поспособствуют развитию нейросетевых алгоритмов стилизации, делая их результаты более выразительными и привлекательными для конечных пользователей.

Авторские декларации

Финансирование

Работа выполнена в рамках государственного задания в сфере научной деятельности Министерства науки и высшего образования РФ на тему "Модели, методы и алгоритмы искусственного интеллекта в задачах экономики для анализа и стилизации многомерных данных, прогнозирования временных рядов и проектирования рекомендательных систем номер проекта FSSW-2023-0004.

Доступность данных и программного кода

Датасет для стилизации изображений:

<https://github.com/victorkitov/style-transfer-dataset>.

Программа для оценки качества изображений:

<https://github.com/EnriFermi/image-evaluation-app>.

Конфликт интересов

Отсутствует.

Вклад авторов

В. В. Китов – постановка задачи, сбор данных, написание и финализация текста статьи.

В. А. Абрамов – разработка математических методов, численные эксперименты, написание текста статьи.

М. К. Ахтырченко – разработка математических методов, численные эксперименты, написание текста статьи.

Список литературы

- [1] *Gonzalez Rafael C.* Digital image processing. — Pearson education india, 2009.
- [2] Controlling perceptual factors in neural style transfer / Leon A Gatys, Alexander S Ecker, Matthias Bethge et al. // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2017. — Pp. 3985–3993.
- [3] Artflow: Unbiased image style transfer via reversible neural flows / Jie An, Siyu Huang, Yibing Song et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2021. — Pp. 862–871.
- [4] Stytr2: Image style transfer with transformers / Yingying Deng, Fan Tang, Weiming Dong et al. // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. — 2022. — Pp. 11326–11336.
- [5] *Gatys Leon A, Ecker Alexander S, Bethge Matthias.* Image style transfer using convolutional neural networks // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2016. — Pp. 2414–2423.
- [6] *Li Chuan, Wand Michael.* Combining markov random fields and convolutional neural networks for image synthesis // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — Pp. 2479–2486.
- [7] *Chen Tian Qi, Schmidt Mark.* Fast patch-based style transfer of arbitrary style // *arXiv preprint arXiv:1612.04337*. — 2016.
- [8] Quality evaluation of arbitrary style transfer: Subjective study and objective metric / Hangwei Chen, Feng Shao, Xiongli Chai et al. // *IEEE Transactions on Circuits and Systems for Video Technology*. — 2022. — Vol. 33, no. 7. — Pp. 3055–3070.
- [9] Progressive growing of gans for improved quality, stability, and variation / Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen // *arXiv preprint arXiv:1710.10196*. — 2017.

- [10] *Rombach Robin, Blattmann Andreas, Lorenz Dominik et al.* High-Resolution Image Synthesis with Latent Diffusion Models. — 2021.
- [11] Adaattn: Revisit attention mechanism in arbitrary neural style transfer / Songhua Liu, Tianwei Lin, Dongliang He et al. // Proceedings of the IEEE/CVF international conference on computer vision. — 2021. — Pp. 6649–6658.
- [12] SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python / Pauli Virtanen, Ralf Gommers, Travis E. Oliphant et al. // *Nature Methods*. — 2020. — Vol. 17. — Pp. 261–272.
- [13] *Kendall M. G.* The Treatment of Ties in Ranking Problems // *Biometrika*. — 1945. — 11. — Vol. 33, no. 3. — Pp. 239–251. — URL: <https://doi.org/10.1093/biomet/33.3.239>.
- [14] Diatom autofocusing in brightfield microscopy: a comparative study / J.L. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, J. Fernandez-Valdivia // Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. — Vol. 3. — 2000. — Pp. 314–317 vol.3.
- [15] *Omkar M. Parkhi Andrea Vedaldi Andrew Zisserman.* Deep Face Recognition // *Proceedings of the British Machine Vision Conference 2015*. — 2015. — Pp. 41.1–41.12.