

АВТОМАТИЧЕСКАЯ СТИЛИЗАЦИЯ ИЗОБРАЖЕНИЙ ПО РАСШИРЕННОЙ КОЛЛЕКЦИИ СТИЛЕЙ

© 2024 г. В. В. Понамарев^{a,*}, В. В. Китов^{a,b,**}

^aМГУ им. М.В. Ломоносова, ф-т ВМК
119991 Москва, ГСП-1, Ленинские горы, д. 1

^bРЭУ им. Г.В. Плеханова, лаборатория искусственного интеллекта
119333 Москва, Стремянный пер., д. 36

*E-mail: valera.pon.vp@gmail.com

**E-mail: v. v.kitov@yandex.ru

Поступила в редакцию: 20.01.2024 г.

После доработки: 24.01.2024 г.

Принята к публикации: 24.01.2024 г.

Стилизация изображений представляет собой актуальную прикладную задачу по автоматической перерисовке исходного изображения (контента) в стиле другого изображения (задающего целевой стиль). Традиционные методы стилизации изображений предоставляют единственный вариант стилизации. Если он пользователя не устраивает, например, в силу возникающих в ходе стилизации артефактов, то ему приходится выбирать другой стиль. В работе предлагается модификация алгоритма стилизации, дающая разнообразные результаты стилизации одним стилем, а также повышающая среднее качество стилизации за счет использования не только стилиевой информации с исходного стилиевого изображения, но и информации с изображений, имеющих похожий стиль.

Ключевые слова: генерация изображений, обработка изображений, нейронные сети

DOI: 10.31857/S0132347424030029, EDN: QAXHSF

ВВЕДЕНИЕ

Алгоритму нейросетевой стилизации изображений (neural style transfer) подаются на вход два изображения — конетное изображение (контент), и стилиевое изображение (стиль). Алгоритм решает задачу автоматической перерисовки контентного изображения в стиле стилиевого изображения. Под стилем понимается цветовая гамма и характерные паттерны рендеринга, такие как мазки кисти художника. Указанная задача актуальна для создания ярких иллюстраций в книгах, на сайтах, в рекламе, дизайне, а также в индустрии развлечений.

Изначально указанная задача была известна под названием non-photorealistic rendering [1–3] и решалась эвристическими методами обработки изображений, подбираемыми под каждый стиль. Нейросетевой перенос стиля, предложенный в работе [4], позволил переносить стиль с произвольного стилиевого изображения-образца, как показано на примере стилизации контентного изображения двумя стилями на рис. 1.

На рис. 1 видно, что качество стилизации существенно зависит от сочетаемости стилей контентного и стилиевого изображений. Если они сочетае-

мы (стиль 2), то результат стилизации получается приемлемым. Если же стили контентного и стилиевого изображений существенно различаются, например, по четкости, как при стилизации первым стилем (см. рис. 1), который более размыт, чем контент, то и стилизация получается неудовлетворительная. Получается, что для создания эффектных стилизаций пользователь вынужден долго перебирать вручную различные стили, пока не найдет стиль, который хорошо сочетается с контентом.

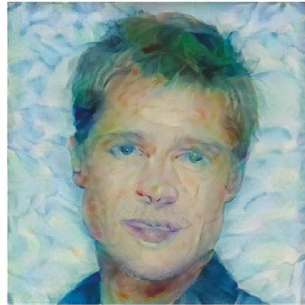
Чтобы упростить процедуру создания качественных стилизаций в работе предлагается производить стилизацию не одним пользовательским стилем, а совокупностью изображений, похожих по стилю на входное пользовательское стилиевое изображение. Для расчета стилиевой похожести предлагается специальная процедура расчета стилиевых характеристик. За счет агрегации стилиевой информации с нескольких изображений повышается устойчивость алгоритма к возможным индивидуальным несоответствиям контентного и стилиевого изображений. Проведенные опросы респондентов показывают, что получаемые стилизации действительно получаются более качественными.



Контент



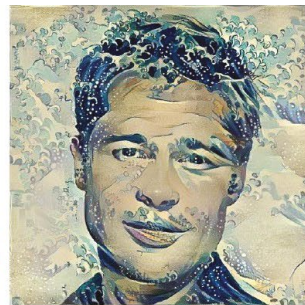
Стиль 1



Стилизация 1



Стиль 2



Стилизация 2

Рис. 1. Применение переноса стиля методом [4]. Первый стиль наложился хуже, чем второй, из-за сильной несочетаемости с контентным изображением (по четкости).

Другим преимуществом предлагаемой модификации является возможность более широко смотреть на стиль. Теперь он не задается одним лишь исходным стилевым изображением, а задан целой совокупностью изображений, имеющих стиль, похожий на целевой. Тем самым, беря разные подмножества из этих изображений, можно получать различные варианты стилизации. Это удобно, если пользователю хотелось бы получить иной вариант стилизации исходным стилем.

Стоит отметить, что базовый подход к стилизации [4], над которым предлагается надстройка, производит оптимизацию в исходном цветовом пространстве стилизуемого изображения, что занимает несколько десятков секунд на видеокарте и неприменимо, например, к стилизации видеопотоков в реальном времени. Много последу-

ющих работ, например [5, 6], было посвящено ускорению стилизации за счет преобразования контентного изображения через специальную нейронную сеть. Поскольку предлагаемый метод представляет собой надстройку над алгоритмом стилизации, то он применим и к ним, а также к другим методам стилизации.

БАЗОВЫЙ АЛГОРИТМ СТИЛИЗАЦИИ ИЗОБРАЖЕНИЙ

Рассмотрим оригинальный подход, указанный в статье [4]. Процесс переноса стиля оперирует со следующими входными данными:

- S — изображение, содержащее желаемый художественный стиль;
- Y — изображение, содержимое которого требуется отобразить в нужном стиле (контент).

Задача алгоритма стилизации — сгенерировать изображение X (стилизацию), на котором объекты контента Y будут изображены в стиле стилизованного изображения S . Стиль задает цветовую гамму и характерные паттерны стилизованного изображения — углы, перебивы между цветами, характерные паттерны, такие как мазки кисти художника.

Для переноса стиля решается оптимизационная задача в пространстве пикселей результирующего стилизованного изображения X :

$$\mathcal{L}_{cont}(X, Y) + \alpha \mathcal{L}_{style}(X, S) \rightarrow \min_X, \quad (1)$$

где потери $\mathcal{L}_{cont}(X, Y)$ штрафуют расхождение контентного изображения и стилизации по смыслу (что именно изображено), а $\mathcal{L}_{style}(X, S)$ штрафуют расхождение стилизованного изображения и стилизации по стилю (как изображено). Гиперпараметр $\alpha > 0$ управляет противоречием между более точной передачей контента (смысла) и более полной передачей стиля.

Расчет контентных и стилевых потерь производится по представлениям изображений X , Y , S на промежуточных слоях сети VGG [7], обученной классифицировать изображения, на выборке ImageNet [8]. При этом для расчета контентных потерь используется один промежуточный слой k , а для расчета стилевых потерь — совокупность более ранних и более поздних слоев, как показано на рис. 2.

Функция потерь для контента:

$$\mathcal{L}_{cont}(X, Y) = \frac{1}{H_k W_k C_k} \sum_{c=1}^{C_k} \sum_{i=1}^{H_k} \sum_{j=1}^{W_k} (X_{cij}^k - Y_{cij}^k)^2,$$

где $X^k, Y^k \in \mathbb{R}^{C_k \times H_k \times W_k}$ — тензоры промежуточных представлений на некотором слое k (гиперпараметр) изображений X и Y после того, как их

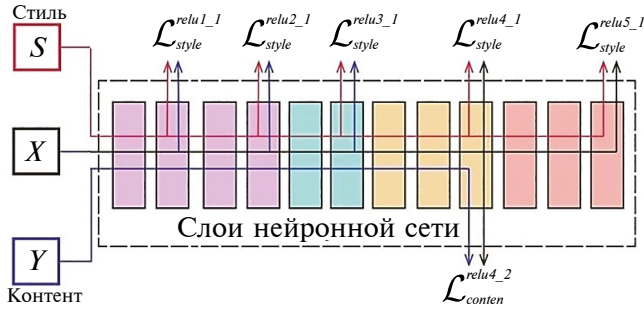


Рис. 2. Используемые при стилизации слои сети VGG.

пропустили через классификационную нейросеть; C_k — число каналов; $H_k \times W_k$ — пространственное разрешение карты признаков. Суть функции заключается в том, что если по смыслу на изображениях X и Y изображены разные объекты, то и их промежуточные представления в сети VGG будут различаться. Штрафуются расхождения именно в промежуточных признаковых представлениях (отвечающих за семантику), а не в исходных RGB-представлениях, поскольку иначе это служило бы слишком сильной привязкой стилизации к исходному контентному изображению и стилизации бы не получилось.

Стилевые потери представляют собой сумму стилизованных потерь по отдельным слоям сети VGG:

$$\mathcal{L}_{style} = \sum_k \mathcal{L}_{style}^k,$$

где

$$\mathcal{L}_{style}^k(X, S) = \frac{1}{C_k^2 H_k^2 W_k^2} \sum_{i=1}^{H_k} \sum_{j=1}^{W_k} \left(G_{ij}^{X^k} - G_{ij}^{S^k} \right)^2. \quad (2)$$

Здесь $G^{Z^k} \in \mathbb{R}^{C_k \times C_k}$ обозначает матрицу Грама, состоящую из скалярных произведений между всевозможными каналами промежуточного представления Z^k в сети VGG на слое k для изображения Z (равного либо стилизованному изображению S либо результату стилизации X):

$$G_{ij}^{Z^k} = \sum_{c=1}^{C_k} \sum_{d=1}^{C_k} Z_{cij}^k Z_{cid}^k. \quad (3)$$

Контент определяет пространственное расположение объектов на изображении, поэтому при расчете контентной функции потерь производится сравнение промежуточных представлений в привязке к пространственным координатам (i, j) .

Стиль же задает общее распределение цветов и более общих паттернов (границ, углов, переливов, мазков кисти и т. д.). Поэтому сначала извлекается это распределение в виде скалярных произведений между каналами (при расчете скалярных

произведений пространственная информация теряется, так как происходит агрегация по всевозможным пространственным координатам i, j), а затем штрафуются уже расхождения распределений признаков между стилизацией и стилем. Такая структура стилизованных потерь приближает стиль, но не приводит к переносу контентной (смысловой) информации со стилизованного изображения.

ПРЕДЛАГАЕМЫЙ ПОДХОД РАСШИРЕННОЙ СТИЛИЗАЦИИ

Поскольку привычка к определенному пользователю изображению стиля может быть слишком жестким ограничением, приводящем к артефактам стилизации, вызванной слабым исходным соответствием контента и стиля (см. рис. 1), предлагается трехэтапный алгоритм стилизации, более устойчивый к индивидуальной несовместимости исходного контентного и стилизованного изображений:

1. Найти для исходного стилизованного изображения наиболее близкие по стилю изображения в широкой базе художественных изображений, таких как Wikiart [9] или Pandora [10] (которая использовалась в работе). Опционально: перекрасить похожие стилизованные изображения в цветовую гамму стиля.

2. Усреднить матрицы Грама по похожим изображениям.

3. Применить алгоритм стилизации (1), заменив в стилизованных потерях (2) матрицы Грама стилизованного изображения на усредненные матрицы Грама по набору изображений, задающих похожий стиль на заданный.

Для поиска похожих изображений по стилю извлекаются вектора поканальных средних для промежуточных представлений в сети VGG. Например, для изображения Z и его промежуточного представления Z^k на слое k получим компоненты C_k -мерного вектора средних следующим образом:

$$v_c^k(Z) = \frac{1}{H_k W_k} \sum_{i=1}^{H_k} \sum_{j=1}^{W_k} Z_{cij}^k, \quad c = \overline{1, C_k}. \quad (4)$$

Далее указанные векторы конкатенируются по более ранним и более поздним слоям VGG, чтобы выравнять соответствие по разным слоям сети. Сравнение изображений по стилю производится сравнением результирующих векторов по Евклидовой норме. Указанное векторное представление содержит информацию о стиле, а не о контенте, поскольку при поканальном усреднении стирается информация о пространственном расположении объектов, а сохраняется только

статистика присутствия тех или иных признаков (цветов и более общих локальных паттернов), которые и характеризуют стиль.

Чтобы повысить соответствие исходному стилю, рекомендуется перекраска наиболее похожих стилизованных изображений на исходный стиль в цвета исходного стиля, используя алгоритм гистограммной перекраски (histogram matching) [11].

Несмотря на то что в следующей секции приводятся результаты запусков для стилизационной модели [4], предлагаемый подход достаточно общий, чтобы быть применимым и для других методов стилизации. Код предложенного подхода доступен на [github](https://github.com/valerapon/Style-transfer-UESC)¹.

СРАВНИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Определим параметры, при которых будет производиться стилизация. Максимальное количество предлагаемых изображений взято равным 3, в качестве метрики близости будем использовать Евклидову метрику, порог близости объектов, при которой будут добавляться похожие изображения, возьмем равным $threshold = 500$. Стилизацию будем проводить по агрегированному (с помощью усреднения матриц Грама) стилю. Сопоставим результаты полученной стилизации со стилизацией с добавленным перекрашиванием найденных стилей в цвет исходного пользовательского стиля.

На рис. 3, 4 показаны некоторые примеры стилизаций с перекрашиванием и без. В обоих случаях видно, что результирующая стилизация немного отлична от исходной, хотя и повторяет ее общие черты. Отсутствие перекрашивания снижает соответствие исходному стилю, но может добавлять яркости результату, что особенно заметно на рис. 3.

Количество стилей в группе является настраиваемым параметром. Используя группы разных размеров, можно получать отличные друг от друга стилизации. Обозначим за k число стилей в группе. На рис. 5, 6 изображены стилизации по группам стилей при разных k . Большее количество изображений позволяет выразительнее передавать детали стилей и контента, но существует условный предел на k , после которого групповая стилизация начинает размываться и терять преимущества относительно стилизации по исходному изображению. Чем больше k , тем менее похожие стили добавляются в группу. Это также зависит от размера всей коллекции изображений. При небольшом объеме уже третий стиль может

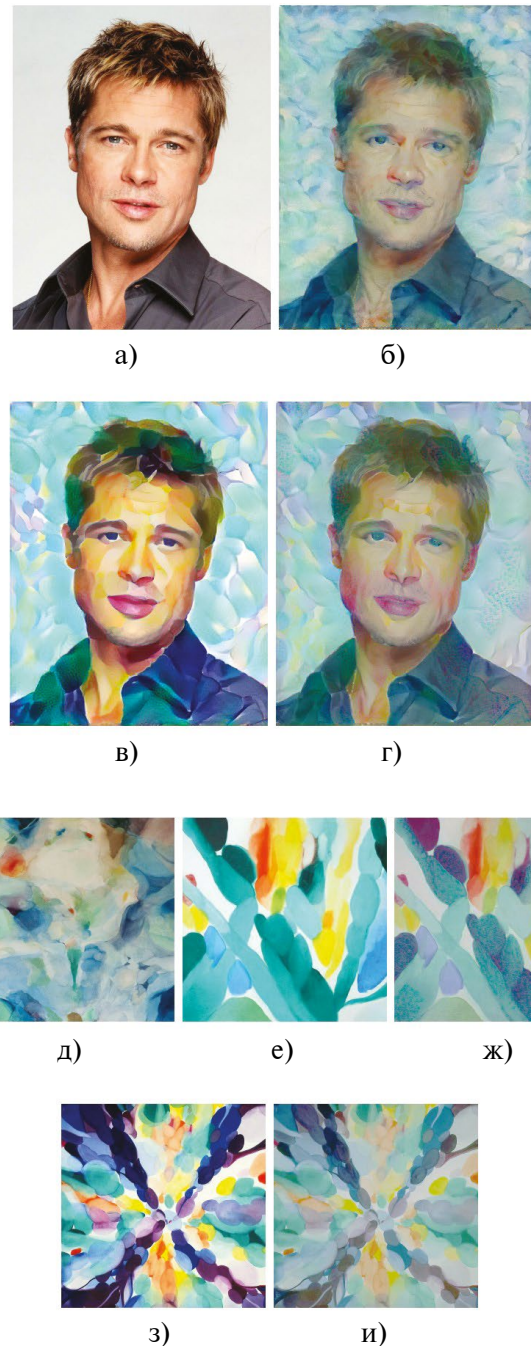


Рис. 3. Пример работы алгоритма стилизации по группе найденных стилей с перекрашиванием исходного стиля: а) контент; б) стандартная стилизация контента а) по стилю д); в) предложенная стилизация по группе найденных стилей е) и з); г) предложенная стилизация по группе стилей с перекрашиванием ж) и и); д) исходный стиль; е) 1-й найденный стиль; ж) стиль е) перекрашенный в д); з) 2-й найденный стиль; и) стиль з) перекрашенный в д).

сильно отличаться от первого. Результаты экспериментов показывают достаточность использования двух-трех дополнительных стилей для повышения качества стилизации.

¹ <https://github.com/valerapon/Style-transfer-UESC>

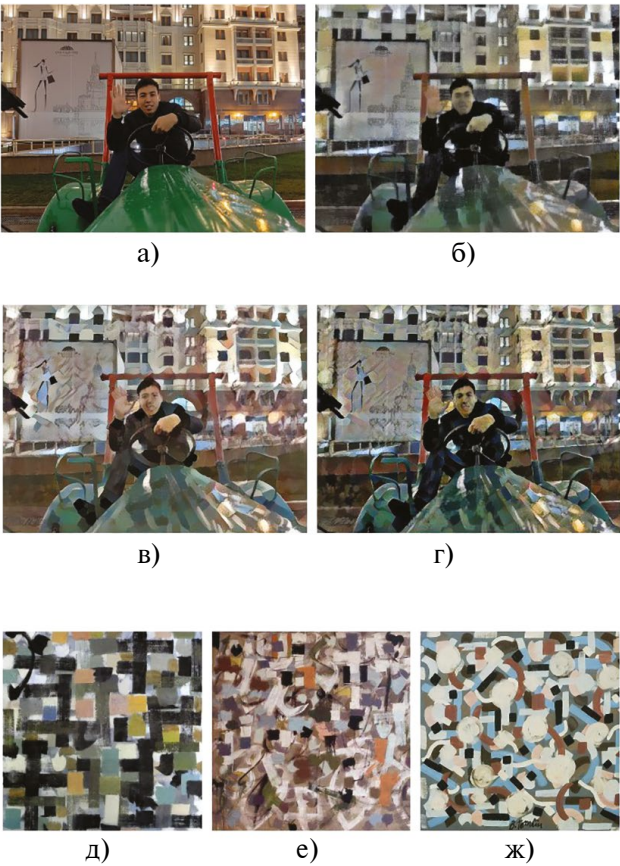


Рис. 4. Пример работы алгоритма стилизации по группе найденных стилей с переносом цветовой схемы исходного стиля: а) контент; (б) стандартная стилизация контента а) по стилю д); в) предложенная стилизация по группе найденных стилей е), ж); г) предложенная стилизация по группе стилей с перекрашиванием е), ж); д) пользовательский стиль; е), ж) похожие на (е) стили.

Таблица 1. Сравнение метрик существующих популярных моделей с предложенным. Значения подсчитывались на тестовом наборе из 1000 пар контент-стиль

Мера качества	Gatys et al. [4]	AdaIN [15]	WCT [16]	Предложенный
Deception rate [14]	0.213	0.117	0.082	0.323
Avg. Gradient Magnitude [12]	160.3	153.9	165.1	177.8
Avg. 90% Quantile FFT [13]	0.488	0.467	0.508	0.541

Таблица 2. Результаты опроса пользователей по сравнению оригинального способа стилизации [4] и предложенного способа стилизации

Число вопросов	Число респондентов	Число ответов	% голосов за новый метод
25	58	1450	71%

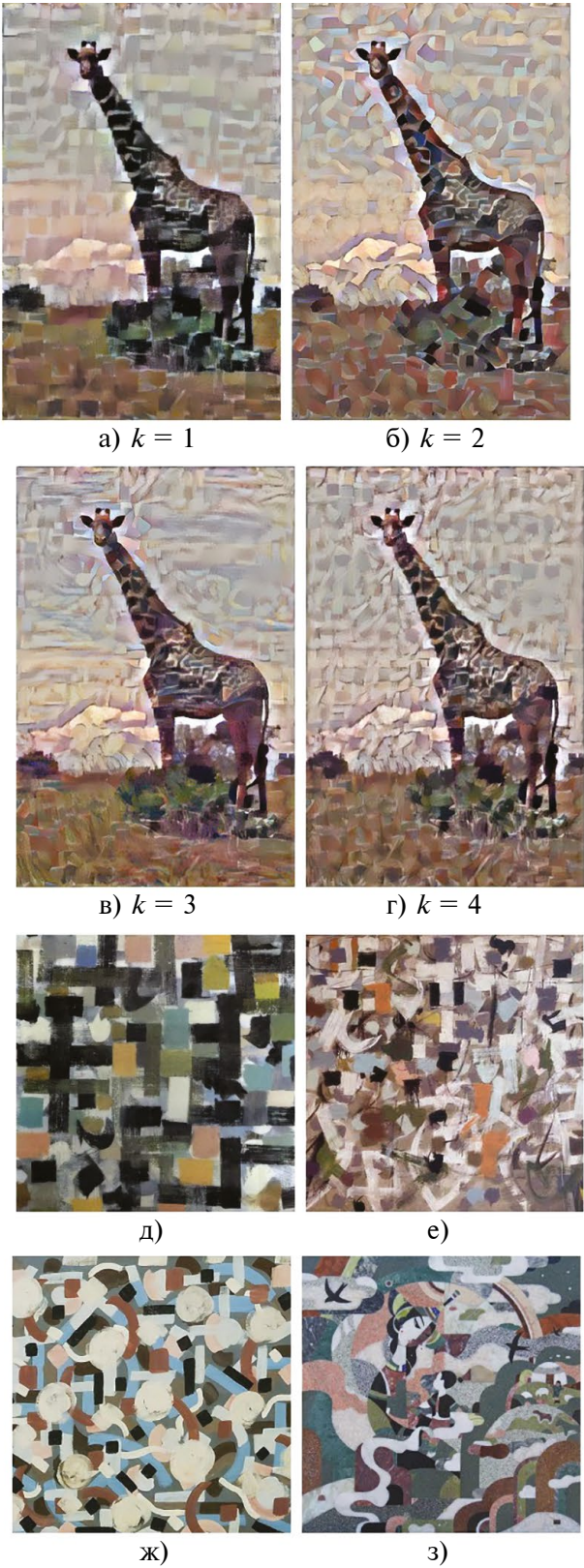


Рис. 5. Работа предложенным алгоритмом стилизации по группе найденных стилей при различном размере группы k : а) стилизация по д), $k = 1$; б) стилизация по д) и е), $k = 2$; в) стилизация по д)–ж), $k = 3$; г) стилизация по д)–з), $k = 4$.



Рис. 6. Работа предложенным алгоритмом стилизации по группе найденных стилей при различном размере группы k : а) стилизация по д), $k = 1$; б) стилизация по д) и е), $k = 2$; в) стилизация по д) – ж), $k = 3$; г) стилизация по д) – з), $k = 4$.

Для численного сравнения стилизаций (см. табл. 1), выполненных разными моделями, были рассмотрены метрики: Average Gradient Magnitude [12] – средняя магнитуд градиентов изображения с окном 3×3 пикселей (оператор Собеля); Average 0.90 Quantile FFT [13] – стилизация делится на 8×8 частей, для каждого блока вычисляются преобразование Фурье и среднее амплитуд, лежащих выше квантили 90%; Deception rate [14] – на основе сети VGG строится классификатор художественных стилей (600 классов), каждый из алгоритмов стилизации тестируется на способность сохранять исходный стиль с помощью построенного классификатора; если он правильно определяет художественный стиль у стилизации, то она считается корректной; оценивается доля корректных стилизаций классификатором на наборе из 1000 пар контент–стиль для каждого алгоритма. По всем мерам качества достигается улучшение предложенного метода по сравнению с базовым [4], использующим лишь один исходный стиль.

Также был проведен опрос среди респондентов, которым для всевозможных пар контент–стиль предлагались на выбор два варианта стилизаций – базовым методом [4] и предложенным (группой похожих стилей с их перекраской под исходный стиль). Респондентам предлагалось выбрать, на их взгляд, более удачный вариант стилизации. Варианты стилизаций каждый раз предлагались в случайном порядке, а респонденты не были ознакомлены с деталями алгоритмов стилизации. Всего участвовало 58 респондентов, каждый из которых сравнивал 25 пар стилизаций. Результаты опросов представлены в табл. 2.

Результат опроса пользователей показал, что пользователи в 71% случаев выбирали новый метод стилизации, что показывает его преимущество по отношению к базовому методу стилизации [4].

ЗАКЛЮЧЕНИЕ

В работе исследовалась задача автоматической стилизации изображений. Было предложено расширить исходное стилевое изображение выборкой похожих по стилю изображений. Для стилевого сопоставления использовались вектора поканальных средних в представлении VGG, которые хорошо себя показали на практике. Стилизация группой похожих стилей, а не только исходным, дает более устойчивый и качественный результат, что показали формальные меры качества и опросы респондентов, чаще предпочитавшие стилизации предложенным методом.

Дополнительным преимуществом метода является возможность изменять группу похожих стилизованных изображений, что позволит пользователю получать различные варианты стилизаций, если первоначальная стилизация пользователя не устраивает.

БЛАГОДАРНОСТЬ

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.

СПИСОК ЛИТЕРАТУРЫ

1. Gooch B., Gooch A. Non-photorealistic rendering. CRC Press, 2001.
2. Strothotte T., Schlechtweg S. Non-photorealistic computer graphics: modeling, rendering, and animation. Morgan Kaufmann, 2002.
3. Rosin P., Collomosse J. Image and video-based artistic stylisation. Springer Science & Business Media, 2012. V. 42.
4. Gatys L.A., Ecker A.S., Bethge M. A neural algorithm of artistic style // arXiv preprint arXiv:1508.06576, 2015.
5. Huang X., Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization // Proceedings of the IEEE international conference on computer vision. 2017. P. 1501–1510.
6. Li Y. et al. Universal style transfer via feature transforms // Advances in neural information processing systems. 2017. V. 30.
7. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // arXiv preprint arXiv:1409.1556, 2014.
8. Deng J. et al. Imagenet: A large-scale hierarchical image database // 2009 IEEE conference on computer vision and pattern recognition. IEEE. 2009. P. 248–255.
9. Phillips F., Mackintosh B. Wiki Art Gallery, Inc.: A case for critical thinking // Issues in Accounting Education. 2011. V. 26. № 3. P. 593–608.
10. Florea C. et al. Pandora: Description of a painting database for art movement recognition with base-lines and perspectives // 2016 24th European Signal Processing Conference (EUSIPCO). IEEE. 2016. P. 918–922.
11. Buzuloiu V. et al. Adaptive-neighborhood histogram equalization of color images // Journal of Electronic Imaging. 2001. V. 10. № 2. P. 445–459.
12. de Villiers J.P. A comparison of image sharpness metrics and real-time sharpening methods with GPU implementations // Proceedings of the 7th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa. 2010. P. 53–62.
13. Hassen R., Wang Z., Salama M.M.A. Image sharpness assessment based on local phase coherence // IEEE Transactions on Image Processing. 2013. V. 22. № 7. P. 2798–2810.
14. Sanakoyeu A. et al. A style-aware content loss for real-time hd style transfer // Proceedings of the European conference on computer vision (ECCV). 2018. P. 698–714.
15. Huang X., Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization // Proceedings of the IEEE international conference on computer vision. 2017. P. 1501–1510.
16. Li Y. et al. Universal style transfer via feature transforms // Advances in neural information processing systems. 2017. V. 30.

AUTOMATIC IMAGE STYLE TRANSFER USING AN AUGMENTED STYLE SET

© 2024 V. V. Ponamaryov^a, V. V. Kitov^b

^aLomonosov Moscow State University, *f*-ty CMC,
Moscow, 119991 Russia

^bPlekhanov Russian University of Economics, Artificial Intelligence Laboratory,
Moscow, 117997 Russia

Image style transfer is an applied task for automatic rendering of the original image (content) in the style of another image (specifying the target style). Traditional image stylization methods provide only a single stylization result. If the user is not satisfied with it due to stylization artifacts, he has to choose a different style. The work proposes a modified stylization algorithm, giving a variety of stylization results, and achieves improved stylization quality by using additional style information from similar styles.

Keywords: image generation, image processing, diversity of generations, neural networks