

НЕЙРОСЕТЕВОЙ ПЕРЕНОС ПРОИЗВОЛЬНОГО СТИЛЯ НА ПОРТРЕТНЫЕ ИЗОБРАЖЕНИЯ С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕЙ С МЕХАНИЗМОМ ВНИМАНИЯ*

С.А. БЕРЕЗИН¹, В.М. ВОЛКОВА²

¹ 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, студент кафедры теоретической и прикладной информатики. E-mail: sergeyberezin123@gmail.com

² 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, доцент кафедры теоретической и прикладной информатики. E-mail: volkova@ami.nstu.ru

Задача переноса произвольного стиля состоит в создании нового, ранее не существовавшего изображения путем комбинирования двух данных изображений – оригинального и стилизового. Оригинальное изображение формирует структуру, основные геометрические линии и формы результирующего изображения, в то время как стилизованное изображение задает цвет и текстуру результата. Слово «произвольный» в данном контексте обозначает отсутствие какого-то одного, заранее выученного стиля. Так, например, сверточные нейронные сети способны переносить новый стиль только после своего переобучения или дообучения на новом объеме данных, не считаются решающими такую задачу, в то время как сети на основе механизма внимания способны производить такую трансформацию без переобучения. Оригинальное изображение может представлять собой, например, фотографию, а стилизованное – картину знаменитого художника. Результирующим изображением в таком случае будет сцена, изображенная на исходной фотографии, выполненная в стилистике данной картины. Современные алгоритмы переноса произвольного стиля позволяют добиться хороших результатов в данной задаче, однако при обработке портретных изображений людей результат работы таких алгоритмов оказывается либо неприемлемым ввиду чрезмерного искажения черт лица, либо слабо выраженным, не несущим характерные черты стилизованного изображения. В этой работе рассматривается подход к решению данной проблемы с использованием комбинированной архитектуры глубоких нейронных сетей с механизмом внимания и сверточной сегментационной сети, осуществляющей перенос стиля с учетом содержимого конкретного сегмента изображения: с ярким преобладанием стиля над формой для фоновой части изображения и с преобладанием содержания над формой в портретной части изображения, содержащей непосредственно изображение лица человека и/или его фигуры.

Ключевые слова: машинное обучение, глубокое обучение, нейронные сети, обработка изображений, перенос стиля, сегментация, сверточные нейронные сети, механизм внимания

* Статья получена 20 ноября 2019 г.

ВВЕДЕНИЕ

Методы машинного обучения в общем и нейронные сети в частности на сегодняшний день находят свое применение в огромном множестве областей [1, 9–15]. Одной из таких областей является обработка изображений. Рассмотрим подробнее задачу переноса стиля, относящуюся к этой области.

Перенос художественного стиля изображения являет собою создание нового изображения, содержащего характерные глобальные и локальные паттерны стилизованного изображения и при этом сохраняющего структуру изображения исходного [1].

Архитектуры на основе сверточных нейронных сетей успешно справляются с переносом одного [2] или нескольких [3] заранее выученных стилей, однако при выходе за рамки заранее заготовленных шаблонов они требуют полного переобучения.

Ранее предпринимались попытки решить эту задачу с использованием техники адаптивной нормализации данных (*adaptive instance normalization*) [4], однако субъективная оценка получаемых результатов была далека от идеальной.

С выходом работ за авторством L. Sheng, Z. Lin, J. Shao, X. Wang [6] и Dae Young Park и Kwang Hee Lee [1] в 2018 году, в которых описывается принципиально новый подход к решению этой задачи, произошел значительный рывок в данной области. Предложенные в архитектуре Avatar-Net [6] и доработанные в архитектуре SANet [1] решения с использованием механизма внимания (*attention mechanism*) позволяют соответствующим образом перестроить характерный паттерн переносимого стиля для каждого участка обрабатываемого изображения с учетом содержания данного участка путем сопоставления отношений, таких как близость (*identity loss*), между контекстным и стилизованным изображениями.

Однако при обработке портретных изображений любым из вышеприведенных методов неизбежно возникает проблема искажения черт лица: те геометрические преобразования, выученные нейросетью, делающие фон похожим по стилю на стилизованное изображение, совершенно искажают форму овала лица, глаз и рта, что приводит к утрате узнаваемости в портрете конкретного человека.

Объектом исследования данной работы стал поиск путей решения данной проблемы, позволяющих сохранить как выразительные геометрические и цветовые преобразования, производимые при переносе стиля, так и узнаваемость черт человеческого лица.

1. МЕТОД

Решение, предлагаемое в данной работе, заключается в разбиении исследуемой задачи на две: перенос стиля на фоновую часть исходного изображения с большим весом формы над содержанием, что позволяет сохранить яркие цветовые и геометрические преобразования, и перенос стиля на лицо и фигуру человека с преобладанием веса содержания над формой. Визуализация решения представлена на рис. 1.

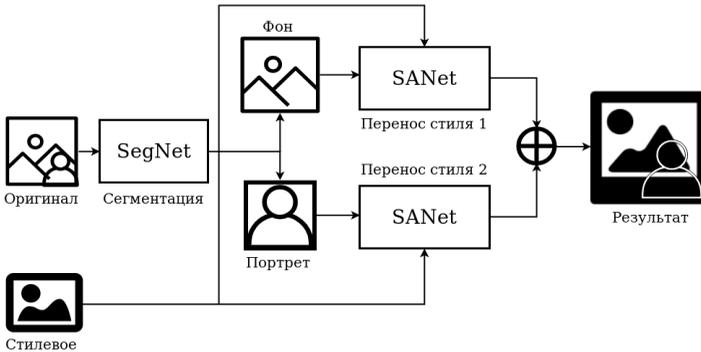


Рис. 1. Схема архитектуры решения

Для разделения изображения на две семантические области: фон и фигуру человека, то есть для решения задачи сегментации, была применена нейросеть архитектуры SegNet [5], обученная на наборе данных Microsoft COCO [7].

Выбор именно этой архитектуры обусловлен малыми временными затратами на обработку кадра (около 40 мс на GPU Nvidia Tesla v100) и удовлетворительными для данной задачи результатами.

Для переноса стиля использовалась упомянутая ранее архитектура SANet (Style-Attentional Network), которая является передовым решением на текущий момент, обеспечивая наиболее качественный результат [1].

Одной из ключевых особенностей этой модели является использование механизма внимания (attention mechanism) для выявления паттернов, характерных для каждого конкретного изображения.

Для варьирования соотношения признаков оригинального изображения к признакам стилового было предложено ввести два весовых коэффициента. Фактически это осуществлялось путем слияния результатов работы двух слоев нейросети с разными коэффициентами:

$$Res = sanet4_1 \cdot W_1 + sanet4_2 \cdot W_2,$$

где коэффициенты W_1 и W_2 определяют степень важности низкоуровневых и высокоуровневых признаков соответственно.

После выделения области с изображением человека составляется бинарная маска, покрывающая эту область. На каждом изображении по полученной маске обрезается фрагмент кадра. Затем производится первый (фоновый) перенос стиля на оригинальное изображение и второй (портретный) перенос стиля. После этого полученные изображения накладываются друг на друга.

2. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Исходная нейросеть SANet была обучена на наборе данных MS-COCO [7] в качестве исходных изображений и на наборе WikiArt [8] в качестве стилевых изображений. Оба набора содержат приблизительно 80 000 изображений.

Целевой функцией была выбрана функция identity loss, направленная на приоритизацию сохранения структуры изображения, а не на изменение стилевых характеристик:

$$L_{identity} = \lambda_{identity1} (||I_{cc} - I_c||_2 + ||I_{ss} - I_s||_2) + \lambda_{identity2} \sum_{i=1}^L (|\phi_i(I_{cc} - I_c)|_2 + \phi_i(||I_{ss} - I_s||_2)),$$

где I_{cc} и I_{ss} обозначают выходное сгенерированное изображение из двух одинаковых исходных и стилевых изображений, I_c и I_s – исходное и стилевое изображения, каждый ϕ_i обозначает слой нейросети, а $\lambda_{identity1}$ и $\lambda_{identity2}$ – гиперпараметры с экспериментально подобранными значениями 1 и 50 соответственно. Как результат, такая функция потерь позволяет сохранять структуру исходного изображения и стилевые особенности переносимого изображения в одно и то же время [1].

Использовался оптимизатор Adam (adaptive moment estimation) [9] с шагом обучения 0.0001 и размером пакета, равным 5. Применялась аугментация обрезанием части изображений.

На рис. 2 и 3 представлен поэтапный результат работы предложенной сети.



Рис. 2. Перенос стиля на портрет. Слева направо: оригинальное изображение, стилевое изображение, результат прямого переноса стиля, результат сегментации, результат переноса стиля на сегментированное изображение, комбинированный результат



Рис. 3. Результат работы алгоритма. Видно соблюдение формы линий человеческого лица при стилизации

ЗАКЛЮЧЕНИЕ

Сочетание всех вышеприведенных методов и техник позволило добиться улучшения качества переноса стиля с произвольного изображения на человеческие портреты в сравнении с исходными подходами. Найденное решение было внедрено в конечный программный комплекс.

СПИСОК ЛИТЕРАТУРЫ

1. *Park D.Y., Lee K.H.* Arbitrary style transfer with style-attentional networks. – URL: <https://arxiv.org/abs/1812.02342> (accessed: 18.11.2019).
2. *Gatys L.A., Ecker A.S., Bethge M.* A neural algorithm of artistic style. – URL: <https://arxiv.org/abs/1508.06576> (accessed: 18.11.2019).
3. *Cui B., Qi C., Wang A.* Multi-style transfer: generalizing fast style transfer to several genres // Stanford University. – 2017. – URL: <http://cs231n.stanford.edu/reports/2017/pdfs/401.pdf> (accessed: 18.11.2019).
4. *Huang X., Belongie S.* Arbitrary style transfer in real-time with adaptive instance normalization. – URL: <https://arxiv.org/abs/1703.06868> (accessed: 18.11.2019).
5. *Badrinarayanan V., Kendall A., Cipolla R.* SegNet: a deep convolutional encoder-decoder architecture for image segmentation. – URL: <https://arxiv.org/abs/1511.00561> (accessed: 18.11.2019).
6. Avatar-Net: multi-scale zero-shot style transfer by feature decoration / L. Sheng, Z. Lin, J. Shao, X. Wang. // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – Salt Lake City, UT, 2018. – P. 8242–8250.
7. Microsoft COCO: Common objects in context / T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Doll'ar, C.L. Zitnick // Computer Vision – ECCV 2014. – Cham: Springer, 2014. – P. 740–755.
8. *Phillips F., Mackintosh B.* Wiki Art Gallery, Inc.: a case for critical thinking // Issues in Accounting Education. – 2011. – Vol. 26 (3). – P. 593–608.
9. *Kingma D.P., Ba J.L.* Adam: a method for stochastic optimization. – URL: <https://arxiv.org/abs/1412.6980v9> (accessed: 18.11.2019).
10. *Хомутов С.О., Серебряков Н.А.* Повышение качества краткосрочного прогнозирования электропотребления группы точек поставки электроэнергии сельхозпроизводителей с помощью инструментов машинного обучения // Научный вестник НГТУ. – 2019. – № 3 (76). – С. 149–168. – DOI: 10.17212/1814-1196-2019-3-149-168.
11. *Гергет О.М., Мецержков Р.В.* Применение информационных методов, нейронных сетей и генетического алгоритма для решения задачи выбора схемы лечения // Научный вестник НГТУ. – 2018. – № 3 (72). – С. 7–20. – DOI: 10.17212/1814-1196-2018-3-7-20.
12. *Воевода А.А., Романников Д.О.* Синтез нейронной сети для решения логико-арифметических задач // Труды СПИИРАН. – 2017. – Вып. 5 (54). – С. 205–223.
13. *Романников Д.О.* О преобразовании сети Петри в нейронную сеть // Сборник научных трудов НГТУ. – 2016. – № 4 (86). – С. 98–103.

14. Дубенко Ю.В., Дышкант Е.Е. Нейросетевой алгоритм выбора методов для прогнозирования временных рядов // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2019. – № 1. – С. 51–60. – DOI: 10.24143/2072-9502-2019-1-51-60.

15. Тимофеев А.В., Дерин О.А. Принципы построения иерархических нейронных сетей для анализа мультиизображений // Труды СПИИРАН. – 2009. – Вып. 10. – С. 160–166.

Березин Сергей Андреевич, студент кафедры теоретической и прикладной информатики Новосибирского государственного технического университета, техник лаборатории бизнес-решений на основе искусственного интеллекта Московского физико-технического института. Основное направление научных исследований – технологии машинного обучения. E-mail: sergeyberezin123@gmail.com

Волкова Виктория Михайловна, доцент кафедры теоретической и прикладной физики Новосибирского государственного технического университета, кандидат технических наук. Основное направление научных исследований – машинное обучение, анализ данных. Имеет более 30 публикаций. E-mail: volkova@ami.nstu.ru

DOI: 10.17212/2307-6879-2019-3-4-96-105

Neural arbitrary style transfer for portrait images using the attention mechanism*

S.A. Berezin¹, V.M. Volkova²

¹ *Novosibirsk State Technical University, 20 Karl Marx Prospekt, Novosibirsk, 630073, Russian Federation, undergraduate of the department of theoretical and applied informatics. E-mail: sergeyberezin123@gmail.com*

² *Novosibirsk State Technical University, 20 Karl Marx Prospekt, Novosibirsk, 630073, Russian Federation, PhD in engineering, associate professor of the department of theoretical and applied informatics. E-mail: volkova@ami.nstu.ru*

Arbitrary style transfer is the task of synthesis of an image that has never been seen before, using two given images: content image and style image. The content image forms the structure, the basic geometric lines and shapes of the resulting image, while the style image sets

* Received 20 November 2019.

the color and texture of the result. The word "arbitrary" in this context means the absence of any one pre-learned style. So, for example, convolutional neural networks capable of transferring a new style only after training or retraining on a new amount of data are not considered to solve such a problem, while networks based on the attention mechanism that are capable of performing such a transformation without retraining - yes. An original image can be, for example, a photograph, and a style image can be a painting of a famous artist. The resulting image in this case will be the scene depicted in the original photograph, made in the style of this picture. Recent arbitrary style transfer algorithms make it possible to achieve good results in this task, however, in processing portrait images of people, the result of such algorithms is either unacceptable due to excessive distortion of facial features, or weakly expressed, not bearing the characteristic features of a style image. In this paper, we consider an approach to solving this problem using the combined architecture of deep neural networks with an attention mechanism that transfers style based on the contents of a particular image segment: with a clear predominance of style over the form for the background part of the image, and with the prevalence of content over the form in the image part containing directly the image of a person.

Keywords: machine learning, deep learning, neural networks, image processing, style transfer, segmentation, convolutional neural networks, attention mechanism

REFERENCES

1. Park D.Y., Lee K.H. *Arbitrary style transfer with style-attentional networks*. Available at: <https://arxiv.org/abs/1812.02342> (accessed 18.11.2019).
2. Gatys L.A., Ecker A.S., Bethge M. *A neural algorithm of artistic style*. Available at: <https://arxiv.org/abs/1508.06576> (accessed 18.11.2019).
3. Cui B., Qi C., Wang A. Multi-style transfer: generalizing fast style transfer to several genres. *Stanford University*, 2017. Available at: <http://cs231n.stanford.edu/reports/2017/pdfs/401.pdf> (accessed 18.11.2019).
4. Huang X., Belongie S. *Arbitrary style transfer in real-time with adaptive instance normalization*. Available at: <https://arxiv.org/abs/1703.06868> (accessed 18.11.2019).
5. Badrinarayanan V., Kendall A., Cipolla R. *SegNet: a deep convolutional encoder-decoder architecture for image segmentation*. Available at: <https://arxiv.org/abs/1511.00561> (accessed 18.11.2019).
6. Sheng L., Lin Z., Shao J., Wang X. Avatar-Net: multi-scale zero-shot style transfer by feature decoration. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 8242–8250.
7. Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollar P., Zitnick C.L. Microsoft COCO: common objects in context. *Computer Vision – ECCV 2014*. Cham, Springer, 2014, pp. 740–755.
8. Phillips F., Mackintosh B. Wiki Art Gallery, Inc.: a case for critical thinking. *Issues in Accounting Education*, 2011, vol. 26 (3), pp. 593–608.

9. Kingma D.P., Ba J.L. *Adam: a method for stochastic optimization*. Available at: <https://arxiv.org/abs/1412.6980v9> (accessed 18.11.2019).
10. Khomutov S.O., Serebryakov N.A. Povyshenie kachestva kratkosrochnogo prognozirovaniya elektropotrebleniya gruppy toчек postavki elektroenergii sel'khozproizvoditelei s pomoshch'yu instrumentov mashinnogo obucheniya [Increasing the quality of short-term load forecasting of the delivery point cluster of agricultural producers with a learning machine]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2019, no. 3 (76), pp. 149–168. DOI: 10.17212/1814-1196-2019-3-149-168.
11. Gerget O.M., Meshcheryakov R.V. Primenenie informatsionnykh metodov, neironnykh setei i geneticheskogo algoritma dlya resheniya zadachi vybora skhemy lecheniya [Applying information methods, neural networks and genetic algorithms for solving the problem of selecting a scheme of treatment]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2018, no. 3 (72), pp. 7–20. DOI: 10.17212/1814-1196-2018-3-7-20.
12. Voevoda A.A., Romannikov D.O. Sintez neironnoi seti dlya resheniya logiko-arifmeticheskikh zadach [Synthesis of neural network for solving logical-arithmetic problems]. *Trudy SPIIRAN – SPIIRAS Proceedings*, 2017, iss. 5 (54), pp. 205–223.
13. Romannikov D.O. O preobrazovanii seti Petri v neironnuui set' [On the transformation of Petri nets in neural network]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2016, no. 4 (86), pp. 98–103.
14. Dubenko Yu.V., Dyshkant E.E. Neurosetevoi algoritm vybora metodov dlya prognozirovaniya vremennykh ryadov [Neural network algorithm for choosing methods of time series forecasting]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika i informatika – Vestnik of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*, 2019, no. 1, pp. 51–60. DOI: 10.24143/2072-9502-2019-1-51-60.
15. Timofeev A.V., Derin O.A. Printsipy postroeniya ierarkhicheskikh neironnykh setei dlya analiza mult'iizobrazhenii [Principles of hierarchical neural networks for analysis of multi-images]. *Trudy SPIIRAN – SPIIRAS Proceedings*, 2009, iss. 10, pp. 160–166.

Для цитирования:

Березин С.А., Волкова В.М. Нейросетевой перенос произвольного стиля на портретные изображения с использованием нейросетей с механизмом внимания // Сборник научных трудов НГТУ. – 2019. – № 3–4 (96). – С. 96–105. – DOI: 10.17212/2307-6879-2019-3-4-96-105.

For citation:

Berezin S.A., Volkova V.M. Neurosetevoi perenos proizvol'nogo stilya na portretnye izobrazheniya s ispol'zovaniem neurosetei s mekhanizmom vnimaniya [Neural arbitrary style transfer for portrait images using the attention mechanism]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2019, no. 3–4 (96), pp. 96–105. DOI: 10.17212/2307-6879-2019-3-4-96-105.