УДК 004.935

© 2020 г. **В.Ю. Ковалев,**
А.Г. Шишкин, д-р физ.-мат. наук
(Московский государственный университет им. М.В. Ломоносова)

РАСПОЗНАВАНИЕ СТИЛЯ ПРОИЗВЕДЕНИЙ ЖИВОПИСИ ПО ИХ ИЗОБРАЖЕНИЯМ С ПОМОЩЬЮ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ

Представлены метод и архитектура нейронной сети для решения задачи классификации стилей произведений живописи по их цифровым изображениям. В результате большого числа проведенных экспериментов по поиску оптимальных гиперпараметров нейронной сети получена итоговая архитектура модели, показавшая точность на 5 близкородственных классах, равную 51,5% и 91% для 5 визуально различающихся стилей.

Ключевые слова: искусственные нейронные сети, глубокое обучение, классификация изображений, набор данных.

DOI: 10.22250/isu.2020.65.74-86

1. Введение

В изобразительном искусстве существует огромное разнообразие стилей, начиная с ренессанса и заканчивая стрит-артом. Во многих случаях они не имеют четко определенных границ, что иногда затрудняет их идентификацию даже экспертами. Поэтому задача автоматического определения стиля картин в настоящее время является весьма актуальной. В целом ее можно рассматривать как составную часть другой задачи, заключающейся в определении авторства произведений живописи по их цифровым изображениям и, возможно, состоящей из ряда таких подзадач, как автоматическое предсказание автора художественного полотна (К. Моне, В. Ван Гог и др.), стиля живописи (импрессионизм, барокко и др.), а также жанра произведения (портрет, пейзаж и т.д.) [1]. Эти три задачи очень сложны из-за большого количества как межклассовых, так и внутриклассовых вариаций: в действительности в одном и том же художественном стиле существуют разные лич-

ные стили, и один и тот же художник может создавать произведения в одном или нескольких различных живописных стилях и жанрах.

Большая часть работ, в которых рассматривается решение указанной задачи, основана на применении классических методов компьютерного зрения. Например, в работах [1, 2] разработаны алгоритмы, основанные на извлечении признаков самого изображения – таких как SIFT (масштабно-инвариантная трансформация признаков) и HOG (гистограмма направленных градиентов) соответственно. В последнем случае признаки, полученные с помощью SIFT, использовались вместе с моделью “мешка слов” (bag of words) для определения одного из восьми авторов картины. В работе [3] представлено сравнительное исследование для решения задачи классификации стилей, в которой сравнивались функции, получающие признаки – такие как SIFT и COLOR SIFT [4], с семантическими признаками, названными Classemes [5] и кодирующими наличие объекта в изображении. Было установлено, что семантико-уровневые признаки значительно превосходят низкоуровневые признаки для этой задачи. Однако оценка была выполнена на небольшом наборе данных для классификации произведений живописи по 7 стилям с 70 цифровыми изображениями для каждого стиля. Авторы [6] также пришли к выводу, что низкоуровневые текстурные и цветовые признаки не эффективны из-за несогласованных цветовых и текстурных узоров, которыми описываются стили в живописи.

Относительно недавно в работе [7] были использованы метрические подходы к обучению модели для поиска взаимосвязей между художниками, основанных на их картинах. Авторы оценивали качество алгоритма по трем метрикам; для оптимизации метрики использовались низкоуровневые признаки, полученные методом HOG. Отметим, что набор данных, используемый в [7], представлен всего лишь 1710 изображениями картин 66 художников.

Основываясь на результатах, полученных при помощи глубоких нейронных сетей, применяемых в качестве инструмента для извлечения характерных признаков в ходе решения различных задач [9, 10], авторы [11] использовали предварительно обученные нейронные сети для работы с небольшим количеством изображений с целью классификации автора картины и ее стиля. В ряде работ [12, 13] были проведены различные эксперименты по разработке архитектуры и обучению сети, а также по совершенствованию существующих сетей для решения задач распознавания стиля и автора живописных полотен. В [12] использовалась предварительно обученная на наборе данных ILSVRC-2012 нейронная сеть, дообученная на наборе дан-

ных, предоставленным *wikipaintings*. Видно, что авторы смогли, используя 20% данных, добиться точности для решения задачи классификации, равной 35% (усредненная точность для всех классов), а используя весь набор данных, – 45%.

В [14, 15] были выполнены исследования, связанные с количеством данных, необходимых для точной настройки сети для задачи классификации стилей произведений живописи. Внутрислойные и межслойные корреляционные признаки в качестве стилевых дескрипторов показали лучшие результаты при классификации по сравнению с признаками сверточных нейронных сетей, в которых используются только полносвязные слои [16, 17].

На рис. 1 приведены примеры изображений, представленные в наборе данных музея «Метрополитен». Можно видеть, что в этом наборе присутствуют не только цифровые изображения предметов живописи, но и фотографии статуй и других предметов искусства.

Рис. 1. Примеры изображений предметов искусства из набора данных МЕТ.

В настоящей работе с использованием технологии глубоких нейронных сетей разработана и реализована модель, позволяющая распознавать стиль картины при ограниченных вычислительных ресурсах. Для определения стиля произведения живописи были использованы методы глубокого обучения с использованием технологий обрезания весов и квантизации для уменьшения размера модели и ускорения вычислений. Следует отметить, что большой объем входных данных (свыше 43000 изображений, относящихся к 5 стилям) был собран вручную из открытых источников, в то время как в аналогичных исследованиях обычно использовались проприетарные базы данных с относительно небольшим числом изображений, что не позволяет адекватно оценить полученные результаты.

В разделе 2 описана модель для решения задачи классификации предметов живописи по пяти различным стилям на основе собранных данных, ее особенности, метод выбора гиперпараметров, а также стратегия и способ

обучения модели. Раздел 3 посвящен формированию выборки для анализа из собранного набора данных и алгоритмам предварительной обработки изображений. В разделе 4 выполнен анализ полученных результатов.

2. Модель

В настоящее время существует большое число различных архитектур нейронных сетей, каждая из которых имеет свои достоинства и недостатки. Основы современных сверточных нейронных сетей были заложены в одной из первых известных работ, посвященных сверточным нейронным сетям, – LeNet-5 [18]. Одними из самых удачных архитектур являются AlexNet и ResNet [19].

В исходной схеме сети AlexNet на вход подается черно-белое изображение размером 32x32 пикселя. Однако возможно использование и цветных изображений. На рис. 2 показана в схематичном виде архитектура сети AlexNet. Она состоит из восьми слоев, первые пять из которых сверточные, а остальные – полносвязные.

Рис. 2. Схематичное изображение архитектуры AlexNet для решения задачи ImageNet.

Обучение AlexNet из-за количества параметров сети происходило на двух GPU параллельно, что позволило сократить время обучения в сравнении с обучением на CPU. Также было показано, что использование функции активации ReLU (Rectified Linear Unit):

$$y(x) = \max(0, x)$$

вместо более традиционных функций сигмоиды

$$y(x) = \frac{1}{1 + e^x}$$

и гиперболического тангенса позволяет снизить количество эпох обучения в

6 раз и в значительной степени уменьшить проблему затухания градиентов, свойственную другим указанным функциям активации.

В AlexNet была применена техника отсева (dropout) [20]. Она заключается в случайном “отключении” нейронов на заданном слое с вероятностью p на каждой эпохе. После обучения сети на стадии распознавания веса слоев, для которых использовалась техника отсева, должны быть умножены на $1/p$. Dropout выступает в роли регуляризатора, не позволяя сети переобучаться на заданном наборе данных. Существует несколько объяснений эффективности этой техники. Первое заключается в том, что отсев позволяет нейронам не полагаться на соседние, а обучаться распознавать более глобальные признаки, что улучшает способность сети к обобщению. Второе объяснение состоит в том, что обучение сети с dropout представляет собой усреднение результатов ансамбля искусственных нейронных сетей, каждая из которых является сетью без некоторого числа нейронов [19]. Таким образом, решение принимает не одна сеть, а ансамбль, в котором каждая сеть обучена по-своему определять разные признаки, тем самым снижая вероятность ошибки итоговой сети.

Другой эффективной архитектурой сверточных сетей является ResNet (Residual Network), разработанная в Microsoft Research [18]. Авторы ResNet увидели, что при увеличении числа слоев качество сети начинает уменьшаться. Для устранения этой проблемы использовались пропускающие соединения. Эти соединения (residual connections) пропускают (рис. 3) один или несколько слоев и выполняют сопоставление идентификаторов.

Рис. 3. Пропускающее соединение архитектуры ResNet.

Их выходы добавляются к выходам сверточных слоев. ResNet относительно легко оптимизировать, помимо того ResNet позволяет увеличивать точность с ростом глубины сети, чего с другими архитектурами добиться сложнее. В последнее время с ростом объема данных и соответственно требований к вычислительным ресурсам получило распространение обучение нейронных сетей не со случайной инициализацией весов, а с использованием весов архитектур, предварительно обученных на большом наборе данных, – таком, например, как ImageNet. Это позволяет использовать уже выделенные сетью при обучении определенные признаки для решения новых задач. Во многих случаях это не только ускоряет процесс обучения, но и улучшает ка-

чество получаемых результатов.

В данной работе мы использовали нейросетевую модель, основанную на архитектуре ResNet. Присутствующие в ней пропускающие-соединения позволяют уменьшить проблему переобучения и градиентного затухания при обучении. Существует несколько типов архитектур ResNet с различным количеством нейронов и слоев (например, – 18, 34, 50, 101 и 152 слоя). Большое количество слоев требует огромного объема данных, чтобы избежать переобучения сети, поэтому мы выбрали в качестве оптимального варианта архитектуру ResNet50 с 50 слоями. На вход нейронной сети подавалось цветное RGB-изображение размером 224x224 пикселя. На выходе получался одномерный вектор размером 528. Этот вектор являлся входным для полносвязного слоя с функцией активации

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$$

и с пятью нейронами на выходе, соответствующими пяти классам каждого стиля.

В принципе, при использовании предобученных нейронных сетей можно настраивать веса только последнего полносвязного слоя. Для небольших объемов данных это позволяет избежать переобучения сверточных слоев. В нашем случае набор данных был достаточно большим, поэтому мы настраивали веса на всех слоях нейронной сети, а не только на последнем.

В качестве функции ошибки использовалась перекрестная энтропия:

$$E = -\sum_i^C t_i \log(f(s)_i),$$

где C – число классов (в нашем случае $C = 5$); t_i – вектор принадлежности объекта к классу i ; $f(s)_i$ – вероятность принадлежности объекта к классу i .

Рис. 4. График перекрестной энтропии для бинарного случая.

В качестве оптимизатора для настройки весов нейронной сети использовался метод ADAM (адаптивный градиентный метод с учетом момента [21]) со скоростью обучения, равной 0.005.

Для улучшения сходимости использовался планировщик (scheduler) [22]. Суть его заключается в том, чтобы согласно определенной зависимости от количества эпох (или количества используемых данных для обучения) изменять значение скорости обучения нейронной сети (обычно она понижается с ростом эпохи/числа просмотренных данных). В данной задаче использовалось уменьшение скорости обучения по косинусному закону (рис. 5), так как он показал наилучшие результаты, в отличие от ступенчатого или уменьшающегося до постоянных значений методов.

Рис. 5. Изменение скорости обучения в зависимости от номера эпохи при использовании планировщика с косинусным законом снижения.

Обучение нейронной сети проводилось в течение 100 эпох. В качестве метрики для оценки качества модели использовалась точность.

3. Данные

В качестве входной информации были использованы данные из набора, представленного на платформе kaggle.com, под названием “Painter by numbers” [23]. Он состоит из изображений из открытого источника wiki-art.com, кроме того, там присутствуют изображения из других открытых баз данных. Для каждого изображения указаны его автор, стиль, жанр и год. К сожалению, эта информация представлена не для каждой картины.

Для решения задачи были оставлены пять классов с наибольшим числом объектов в каждом. Заметим, что классы, которые использовались в дальнейшем, содержат достаточно похожие между собой изображения. Например, поставленная задача в значительной мере упростилась бы при использовании для анализа изображений, относящихся к таким сильно отличающимся стилям, как авангардизм, кубизм, поп-арт, реализм и пуантилизм. В исходном наборе данных было распределение изображений по классам:

- 1) импрессионизм – 10643 (24.5 %);
- 2) реализм – 10523 (24.3%);
- 3) романтизм – 9285 (21%);
- 4) экспрессионизм – 7013 (16.2%);
- 5) постимпрессионизм – 5778 (13.3%).

Для обучения было использовано 80%, а для тестирования – 20% всех данных, с сохранением распределения имеющихся классов.

Отметим, что при использовании предобученных нейронных сетей важно, чтобы размер входных данных и тех, на которых нейронная сеть обучалась заранее, совпадал. В связи с этим размер изображений был изменен до 224x224 пикселей. Помимо того, при исследовании были применены различные типы аугментации с целью генерализации выборки и повышения устойчивости обучающейся нейронной сети к шуму. В данной работе использовались следующие преобразования в указанном порядке:

- 1) случайное аффинное преобразование с поворотом изображения до 10° и случайным увеличением изображения в 1.1 – 1.3 раза;
- 2) вырезание случайного участка изображения размером 224x224 пикселей;
- 3) случайное отражение по вертикали;
- 4) небольшое изменение яркости, контраста, насыщенности цвета, тона;
- 5) нормализация.

Суть нормализации заключается в приведении значений пикселей к диапазону, используемому в наборе данных ImageNet. Применение нормализации в ImageNet позволяло бороться с переобучением. Вследствие того, что выделенные нейросетью признаки для нормализованных и ненормализованных изображений могут не совпадать, в данной работе мы использовали нормализацию. Примеры изображений, а также результаты аугментации представлены на рис. 6 и 7.

Рис. 6. Примеры изображений из выборки.

Рис. 7. Изображения до и после использования аугментаций.

4. Результаты

Процесс обучения описанной выше нейросети проводился в течение 100 эпох. Изменения точности и величины функции ошибки в зависимости от номера эпохи обучения представлены на рис. 8.

Рис. 8. Зависимости точности и функции ошибки (перекрестной энтропии) для обучающего и валидационного набора данных от номера эпохи обучения модели.

Из рисунка следует, что к 80 эпохе как значение точности, так и функции ошибки на валидационном множестве выходят на плато, что говорит о том, что модель закончила обучение. Также можно видеть, что точность на обучающем и валидационном множествах близка, это позволяет сделать вывод, что модель не переобучилась.

Для увеличения скорости работы нейронной сети использовалась техника обрезания весов согласно их величине. Она заключается в устранении связи между нейронами, связанными весовыми коэффициентами, имеющими малое значение. Это позволило увеличить скорость работы на 13% при уменьшении качества всего на 2%.

На рис. 9 представлена матрица ошибок для итоговой модели. Различные стили живописи имеют общие черты – такие как цвета, композиции и текстуры, а также общие объекты.

Рис. 9. Матрица ошибок для итоговой модели.

Таким образом, неверная классификация между близкородственными стилями встречается довольно часто (рис. 10 и 11).

Рис. 10. Примеры изображений из валидационного набора, которые классифицировались нейронной сетью неправильно. Стиль изображения (а) определен нейронной сетью как постимпрессионизм, но настоящий класс – импрессионизм. Стиль (б) – наоборот.

Рис. 11. Примеры изображений из валидационного набора, которые классифицировались нейронной сетью неправильно. Стиль изображения (а) определен нейронной сетью как импрессионизм, но настоящий класс – постимпрессионизм. Стиль (б) – наоборот.

Как видим, наилучшие результаты у модели получаются для импрессионизма. Ошибки неправильной классификации, как правило, встречаются значительно чаще среди тесно связанных групп стилей – таких как постимпрессионизм и импрессионизм или романтизм и реализм, отражая схожесть этих направлений и наличие более сложных признаков их различия. Общая точность модели со всеми улучшениями и после использования обрезки весов составила 51.51%.

Если же обучить нейронную сеть для классификации визуально различающихся стилей – таких как: кубизм, поп-арт, реализм и пуантилизм, то результаты (рис. 12) будут заметно лучше – средняя точность для классов – 91%.

Рис. 12. Матрица ошибок для визуально сильно различающихся классов.

Заключение

В данной работе разработаны модель и архитектура глубокой нейронной сети для классификации произведений искусства по пяти стилям: импрессионизм, романтизм, реализм, экспрессионизм и постимпрессионизм. Следует отметить, что для классификации были выбраны стили, весьма схожие между собой, что в значительной мере осложняло поставленную задачу. Во многих случаях даже люди, профессионально занимающиеся историей изобразительного искусства, не могут правильно определить тот или иной родственный стиль живописи в конкретном изображении.

Были использованы различные способы улучшения качества классификации: использование аугментаций, планировщиков, подбор параметров оптимизатора, выбор функции ошибки. Также были выполнены поиск и пре-

доработка существующих наборов данных. С целью уменьшить вероятность переобучения и увеличить вычислительную производительность было произведено обрезание весов нейронной сети. Итоговая точность модели для классификации картины по пяти классам составила 51.5%.

Результаты данной работы могут быть использованы для дальнейших исследований в этой области и решения таких задач как определение подлинности картины, установление стиля и жанра произведения живописи из большого числа различных классов и др.

ЛИТЕРАТУРА

1. *Anwer R.M., Khan F.S., Weijer J., Laaksonen J.* Combining holistic and part-based deep representations for computational painting categorization // Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval. – 2016. – P. 339-342.
2. *Lowe D.G.* Distinctive image features from scale-invariant keypoints // International Journal Computer Vision. – 2004. – Vol. 60. – P. 91-110.
3. *Dalal N., Triggs B.* Histograms of oriented gradients for human detection // International Conference on Computer Vision & Pattern Recognition. – 2004. – Vol. 2. – P. 886-893.
4. *Arora R.S., Elgammal A.M.* Towards automated classification of fine-art painting style: A comparative study // Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). – 2012. – P. 3541-3544.
5. *Abdel-Hakim A.E., Farag A.A.* CSIFT: A SIFT descriptor with color invariant characteristics // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). – 2006. – P. 1978-1983.
6. *Torresani L., Szummer M., Fitzgibbon A.* Efficient object category recognition using classemes // ECCV'10: Proceedings of the 11th European conference on Computer vision: Part I. – 2010. – P. 776-789.
7. *Carneiro G., da Silva N.P., Del Bue A., Costeira J.P.* Artistic image classification: An analysis on the printart database // European Conference on Computer Vision, Springer. – 2012. – P. 143-157.
8. *Khan F.S., Beigpour S., Van de Weijer J., Felsberg M.* Painting-91: a large scale database for computational painting categorization // Machine vision and applications 25. – 2014. – Vol. 6. – P. 1385-1397.
9. *Razavian A.S., Azizpour H., Sullivan J., Carlsson S.* CNN features off-the-shelf: an astounding baseline for recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. – 2014. – P. 806-813.
10. *Bianco S., Mazzini D., Pau D., Schettini R.* Local detectors and compact descriptors for visual search: a quantitative comparison // Digital Signal Processing 44. – 2015. – P. 1-13.
11. *Peng K.-C., Chen T.* A framework of extracting multi-scale features using multiple convolutional neural networks // 2015 IEEE International Conference on Multimedia and Expo (ICME). – 2015. – P. 1-6.
12. *Hentschel C., Wiradarma T.P., Sack H.* Fine tuning CNNs with scarce training data adapting imagenet to art epoch classification // Image Processing (ICIP), 2016 IEEE International Conference. – 2016. – P. 3693-3697.

13. *Krizhevsky A., Sutskever I., Hinton G.E.* Imagenet classification with deep convolutional neural networks // Advances in neural information processing systems. – 2012. – Vol. 25. –P. 1097-1105.
14. *Banerji S., Sinha A.* Painting classification using a pre-trained convolutional neural network // International Conference on Computer Vision, Graphics, and Image processing, Springer. – 2016. – P. 168-179.
15. *Tan W.R., Chan C.S., Aguirre H.E., Tanaka K.* Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification // Image Processing (ICIP), 2016 IEEE International Conference on, IEEE. – 2016. – P. 3703-3707.
16. *Chu W.-T., Wu Y.-L.* Deep correlation features for image style classification // Proceedings of the 2016 ACM on Multimedia Conference, ACM. – 2016. – P. 402-406.
17. *Chu W.-T., Wu Y.-L.* Image style classification based on learnt deep correlation features // IEEE Transactions on Multimedia. – 2018. – Vol. 20. – P. 2491-2502.
18. *LeCun Y., Bottou L., Bengio Y., Haffner P.* Gradient-Based Learning Applied to Document Recognition // Proceedings of the IEEE. – 1998. – Vol. 86. – P. 2278-2324.
19. *Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R.* Dropout: a simple way to prevent neural networks from overfitting // Journal of Machine Learning Research. – 2014. – Vol. 15. – P. 1929-1958.
20. *He K., Zhang X., Ren S., Sun J.* Deep residual learning for image recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2015. – P. 770-778.
21. *Kingma D.P., Lei Ba J.* Adam: A method for stochastic optimization. – 2014. arXiv: 1412.6980v9.
22. *Loshchilov I., Hutter F.* SGDR: stochastic gradient descent with warm restarts // Conference: ICLR 2017 (5th International Conference on Learning Representations). – 2016.
23. Painter by number competition: <https://www.kaggle.com/c/painter-by-numbers>.

Статья представлена к публикации членом редколлегии О.С. Амосовым.

E-mail:

Ковалев Валентин Юрьевич – kovalev.vyu@gmail.com;

Шишкин Алексей Геннадиевич – shishkin@cs.msu.ru.