

Применение методов глубокого обучения для оценки степени коммерческой ценности визуальных объектов

В.Г. Ефремов¹, Н.Г. Ефремов¹, Е.П. Тетерин², П.Е. Тетерин³, В.В. Гансовский¹

¹Независимый исследователь;

²Ковровская государственная технологическая академия им. В.А. Дегтярева,
Владимирская обл., г. Ковров, Россия;

³Национальный исследовательский ядерный университет «МИФИ», г. Москва, Россия

Аннотация

Рассмотрена возможность применения сверточной нейронной сети для оценки коммерческой ценности цифровых изображений. Исследовалось влияние на обучение нейронной сети различных условий подготовки образцов, алгоритмов оптимизаторов, количества пикселей в образцах, размеров обучающей выборки, цветовых схем, качества сжатия и других фотометрических параметров. Показано, что благодаря предложенной предварительной подготовке данных, оптимальному выбору архитектуры и гиперпараметров нейросети удалось добиться точности классификации не менее 98 %.

Ключевые слова: глубокое обучение, нейросети, анализ изображений.

Цитирование: Ефремов, В.Г. Применение методов глубокого обучения для оценки степени коммерческой ценности визуальных объектов / В.Г. Ефремов, Н.Г. Ефремов, Е.П. Тетерин, П.Е. Тетерин, В.В. Гансовский // Компьютерная оптика. – 2020. – Т. 44, № 1. – С. 127-132. – DOI: 10.18287/2412-6179-CO-515.

Citation: Efremtsev VG, Efremtsev NG, Teterin EP, Teterin PE, Gantsovsky VV. Deep learning application for box-office evaluation of images. Computer Optics 2020; 44(1): 127-132. DOI: 10.18287/2412-6179-CO-515.

Введение

С развитием медиапространства и социальных сетей всё большее внимание уделяется составлению рейтингов контента сети Интернет [1, 2]. Существует множество работ, посвящённых оценке популярности изображений и видео с помощью современных программных средств. Весьма распространённым является направление оценки параметров качества изображений, результатом которого может стать как отбор наиболее привлекательных для среднего пользователя изображений, так и создание автоматизированных фильтров для обработки изображений с использованием оптимальных параметров, полученных в результате обучения [3–8].

Известен ряд работ по обучению нейронных сетей (НС) для отбора изображений по их эстетической привлекательности. Для этого используются статистические параметры изображения по различным каналам (RGB, HSV), цветовые гистограммы, гистограммы градаций серого, гистограммы ориентированных градиентов, комплексные векторные параметры (например, вектор Фишера), анализируются различные неоднородности и специфические области изображений с помощью SIFT и т.д. Изображению присваивается рейтинг, представляющий собой суперпозицию большого количества различных параметров [4–13], каждый из которых обладает своим собственным весовым коэффициентом [4, 9]. Важным моментом является учёт «шумовых» характеристик в рассматриваемых распределениях, а также способ их отсеечения.

Для выработки чётких критериев отсеечения и уточнения результатов выборки сужают по тематическим направлениям, представленным в виде тэгов (*tags*) [10–11]. Причём тэги могут присваиваться, исходя из анализа самого изображения с помощью алгоритмов распознавания [14, 15]. Такое уточнение может позволить уменьшить входное разрешение исследуемых образцов, что важно и удобно при анализе изображений на различных Интернет-ресурсах, использующих т.н. превью (*preview*) – уменьшенные для предпросмотра изображения. Глубокое обучение для оценки коммерческой составляющей продукции также используется в различных отраслях. В случае с изображениями обучение НС может происходить с помощью анализа самих изображений, однако в ряде случаев для обучения требуется вводить набор параметров или атрибутов коммерческого продукта, не всегда очевидным образом связанных с его изначальными характеристиками. Например, при оценке популярности фильмов при обучении используется аннотация, написанная зачастую на самой ранней стадии съёмки фильма [16], и даже пользовательские обзоры, а не только оценки [17]. В случае с предсказанием коммерческого успеха одежды тип магазина, где исследуемая одежда продается, входит как параметр со значительным весовым коэффициентом [18]. Различные параметры могут входить в рейтинг с различными весовыми коэффициентами, поэтому для выявления этих фактов при обучении, как правило, необходима максимально большая выборка. Параметры для обучения могут как являться, так и не яв-

ляться частью неотъемлемых характеристик товара для исследуемых товарных категорий.

Таким образом, решающее влияние на эффективность обучения и, как следствие, его применение для анализа конкретных характеристик зависит от сочетания подбора обучающей выборки, исследуемой выборки, выбора интересующих (значимых) параметров и области применимости модели.

Учитывая субъективность оценки эстетический привлекательности тех или иных изображений, можно сформулировать задачу изучения возможности применения алгоритмов глубокого обучения для разделения изображений на классы с точки зрения их коммерческой ценности. Для удобства выбрана бинарная модель с определённым порогом продаж. Коммерческая ценность определяется количеством продаж в фотобанках, которые фиксируют и показывают эту статистику.

Существует ряд коммерческих проектов, направленных на оценку эстетической ценности изображений и видео, некоторые из которых доступны онлайн [19, 20]. К сожалению, для таких решений, как правило, отсутствуют описания архитектуры используемых ими нейросетей. Кроме того, коммерческая ценность и эстетический рейтинг изображений на основе, например, результатов, полученных из социальных сетей, могут не совпадать.

Авторам данной статьи не удалось заручиться поддержкой администраций ряда фотобанков для проведения широкомасштабного исследования и обучения по выборке с большим количеством изображений и разных авторов. Тем более интересным с точки зрения реального применения является создание метода, позволяющего классифицировать изображения с невысоким разрешением и по небольшой выборке работ одного автора.

Практическая задача такого метода состоит в следующем: обученная нейронная сеть (НС) может дать конкретную оценку коммерческого потенциала, весьма субъективного и неочевидного критерия, на который влияют много неучтённых факторов, что позволит на основе конкретных данных корректировать маркетинговые стратегии и технологии продаж.

Для анализа коммерческой ценности фотографий была сформирована выборка из 400 снимков, из которых 200 снимков хорошо продавались (имели более 200 продаж на фотостоках), а 200 снимков плохо, с максимальным размером 600×400 пикселей, цветовой моделью RGB, форматом файлов JPEG и JPEG2000. Уровень коммерческой ценности определялся автором фотографий.

В результате правильного выбора архитектуры, алгоритмов обучения и гиперпараметров НС удалось разделить изображения по двум указанным выше классам с точностью не менее 98% в рамках столь малой выборки.

Предварительная подготовка изображений

Полученные изображения имели различные параметры, и поэтому на первом этапе обработки они были приведены к одному размеру, цветовой модели и формату файлов. Использовались программные возможности MATLAB и средства пакетной обработки изображений программы Photoshop.

После получения положительных результатов уверенного разделения исходных фотографий на два класса было изучено, как на точность классификации влияют технические параметры снимка: размеры, цветовая модель, качество сжатия JPEG и другие фотометрические параметры. Для этого анализировались снимки различных размеров: 400×600, 200×300, 100×150, 67×100, 50×75, 40×60, 33×50, 27×40, 20×30, 17×25, 13×20, 10×15, 8×12, 7×10, 4×6 пикселей – при нормальных значениях яркости, насыщенности и максимальном качестве JPEG. Также для изображения размером 100×150 пикселей была исследована классификация при различных уровнях яркости (0, 0,25, 0,50, 0,75, 1,00), насыщенности (0, 0,25, 0,50, 0,75, 1,00), цветового тона (0, 0,25, 0,50, 0,75, 1,00), качества сжатия JPEG (0, 3, 6, 9, 12). Следует отметить, что изменение величины яркости, насыщенности или цветового тона происходило при неизменности других параметров. Для изображения размером 400×600 изучалась классификация четырёх цветовых пространств: CMYK, LAB, HSV, GRAY. Для цветовых пространств CMYK и LAB использовался формат файла JPEG2000.

Условия обучения НС

В качестве функции потерь была использована кросс-энтропия, которая часто применяется при обучении НС, т.к. даёт очень большие градиентные значения, что особенно ценно для градиентного спуска [22]. В MATLAB эта функция потерь представлена в виде [21]:

$$J = \sum_i^M \{-d_i \ln(y_i) - (1 - d_i) \ln(1 - y_i)\},$$

где y_i – выход НС, d_i – значения меток, M – количество образцов.

Обучение проводилось при различном соотношении количества фото, которые случайным образом отбирались для обучения и тестирования НС. Эти соотношения равнялись: 0,9, 0,8, 0,7, 0,6, 0,5 и 0,2.

В связи с тем, что при запуске НС на обучение каждый раз заново инициализируются начальные значения весов и смещений, а также распределение фото для обучения и тестирования определялось случайным образом, результаты каждого расчёта немного отличаются друг от друга. Поэтому все расчёты проводились при десятикратной повторности и затем усреднялись. Результаты усреднения точности классификации при десятикратной повторности отлича-

лись от результатов стократной повторности не более половины процента.

Для оптимального выбора архитектуры, алгоритмов обучения и гиперпараметров системы глубокого обучения рассчитывалась точность классификации, а также анализировались образцы, неправильно отнесённые к 1-му или 2-му классам на этапах обучения и тестирования (рис. 1).

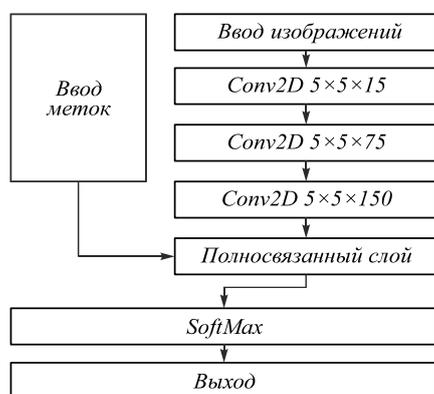


Рис. 1. Архитектура свёрточной НС, используемой для классификации изображений

Предварительная подготовка данных и программирование свёрточной нейронной сети осуществлялось посредством математического пакета MATLAB, Deep Learning Toolbox.

Подготовка данных

Первоначально исходные фотографии, согласно условиям работы свёрточной сети с изображениями, были помещены в две папки, которые имели условные названия GOOD и TECH. В связи с ограниченным количеством изображений для повышения точности классификации и избегания переобучения использовалась предусмотренная в MATLAB операция генерации дополнительных данных путём их случайной трансформации. Наличие функции *imageDataAugmenter* позволило проводить каждый цикл обучения на относительно новых данных. Порождённые данные не сохранялись и тем самым оставляли исходные данные без изменений. Для наших изображений наиболее эффективным было: поворот относительно центра изображения в пределах от -5 до $+5$ градусов, сдвиг вправо и влево не более 5 пикселей, смещение по горизонтали и вертикали в интервале от -5 до $+5$ пикселей. Применение оператора порождения новых данных позволило повысить точность классификации с 75% до 99%, что явилось существенным вкладом в достижение высокой точности. Дополнительные трансформации типа отражения, поворота вокруг вертикальной и горизонтальной оси практически не влияли на точность классификации.

Программное обеспечение

НС состояла из слоя ввода изображения, трёх свёрточных блоков, полносвязанного слоя, слоя при-

менения функции *softmax* и выходного слоя непосредственно классификации.

Первый слой обеспечивал ввод матрицы, соответствующей изображению, а также нормализацию данных. Каждый из трёх свёрточных блоков состоял из 5 слоёв: двухмерного свёрточного слоя, слоя пакетной (*batch*) нормализации, *relu* слоя, слоя нормализации *crossChannel* и слоя *maxPooling*. Наша сеть для первого слоя имела 15 фильтров размером 5×5 пикселей каждый. Для последующих свёрточных слоёв размер фильтра оставался постоянным, а число фильтров для второго и третьего слоёв равнялось соответственно 75 и 150. Шаг сдвига фильтра равнялся одному пикселю. Дополнение пространственных размеров для всех свёрточных слоёв также равно 1 пикселю. Слой пакетной нормализации, как и другие слои нормализации, обеспечивает повышение скорости обучения сети и понижение влияния больших значений весов. Слой функции активации *relu*, стандартный практически для всех свёрточных сетей, необходим для отсекаания отрицательных значений. Выбор максимального значения веса для каждого фильтра (слой *maxPooling*) не только значительно понижает размер матрицы весов, но и способствует выделению более информативных признаков.

Полносвязанный слой, слой *softmax* и выходной слой обеспечивают необходимые преобразования значений весов и смещений, получаемых от третьего свёрточного блока, и вычисляют уровень кросс-энтропийных потерь для осуществления классификации. В полносвязанном слое указывается количество классов, равное двум.

Подготовка к обучению сети

После определения архитектуры нейронной сети были подобраны параметры, определяющие процесс обучения. В качестве алгоритма оптимизации наилучшие значения точности классификации показал метод адаптивной оценки момента ADAM [22]. Для работы оптимизатора необходимо задать размер мини-партии (подмножества обучающей матрицы). Максимальное количество эпох не превышало 75. Хорошие результаты классификации были получены при *MiniBatchSize* = 30. В качестве начальной скорости обучения была выбрана величина 0,0001, которая через 25 эпох понижалась в 10 раз. Для понижения вероятности переобучения значение параметра регуляризации задавалось равным 0,001. Приведённые значения гиперпараметров получены методом сеточного поиска [23]. Типичные графики процесса обучения приведены на рис. 2 и 3. На этапе подготовки сети к обучению также фиксировалось подключение графического процессора.

Оценка качества работы нейронной сети проводилась на тестовой выборке, т.е. на образцах, которые не участвовали в обучении. Расчёт точности классификации проводился по формуле [23]:

$$A = 100 \cdot \frac{n_T}{N_{test}}$$

где A – точность классификации в процентах; n_T – число снимков, правильно отнесенных к тому или иному классу; N_{test} – количество снимков в тестовой выборке.

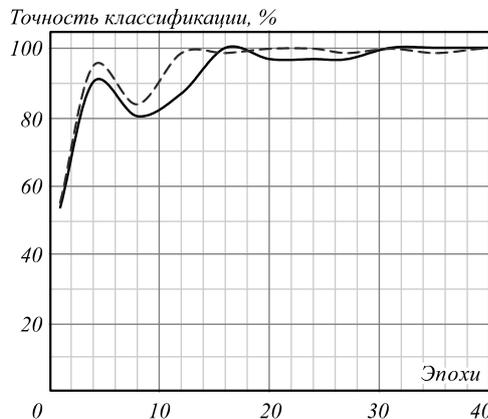


Рис. 2. Изменение точности в зависимости от числа эпох. Сплошной линией показана точность для обучения, пунктирной – для тестирования

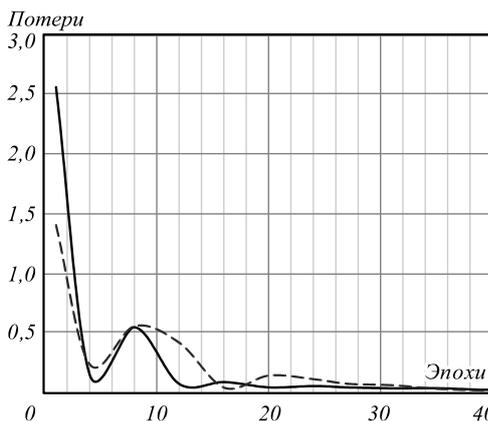


Рис. 3. Изменение функции потерь во время обучения. Сплошной линией показаны потери для обучения, пунктирной – для тестирования

Обучение и тестирование НС проходило на компьютере следующей конфигурации: Intel i7-7700 3,6 GHz, 32 GB DDR4, SSD Samsung 970 Pro 512 GB, Win 10 x64. Для ускорения расчётов использовалась графическая карта NVIDIA GTX 1080Ti, фреймбуфер 11 GB GDDR5X, 3584 NVIDIA CUDA Cores, архитектура GPU Pascal. Среднее время одной эпохи обучения в зависимости от размеров снимков изменялось от 10 до 75 секунд.

Результаты

Значения точности приведены в процентах по результатам классификации фотографий тестовой выборки. На рис. 4 и 5 представлены фотографии, относящиеся к разным классам, с точки зрения коммерческой успешности, которые использовались для обучения. С учётом композиционной схожести изобра-

жений выбор покупателей неочевиден. Интересно, что онлайн-сервис Everypixel aesthetics [20] присвоил оценку указанным изображениям противоположно фактическим значениям продаж, а сервис [19] произвёл распознавание образов и присвоил каждому из них более высокую оценку в случае рис. 3, однако интегральная оценка в данном сервисе недоступна.



Рис. 4. Пример коммерчески успешной фотографии (GOOD)



Рис. 5. Пример фотографии, не являющейся коммерчески успешной (TECH)

Значения насыщенности, яркости и сдвига цветового тона приведены в относительных единицах. За единицу принято соответствующее значение конкретного снимка. Значения компрессии JPEG соответствуют условным единицам Photoshop. Размеры снимков в таблицах представлены в пикселях, первое число отражает число строк, второе – число столбцов.

Табл. 1. Влияние на точность классификации соотношения количества фото в выборках для обучения и тестирования

Соотношение	0,9	0,8	0,7	0,6	0,5	0,25
Точность	100,0	100,0	99,6	99,7	99,5	99,3

Таким образом, даже при небольшой выборке НС показывает высокую точность классификации.

Табл. 2. Влияние на точность классификации цветовой модели

Цветовая модель	RGB	CMYK	LAB	HSV	GRAY
Точность	100,0	100,0	99,5	99,4	99,4

Изменение цветовой модели не влияет на точность классификации.

Также оценивалось влияние на точность классификации компоненты цветовой модели HSV.

Табл. 3. Влияние на точность классификации насыщенности цвета (Saturation)

Насыщенность	0	0,25	0,50	0,75	1,00
Точность	99,2	99,5	99,5	99,5	100,0

Точность классификации не изменяется при изменении насыщенности цвета. Результат при насыщенности цвета, равный нулю, соответствует результату для цветовой модели GRAY.

Табл. 4. Влияние на точность классификации яркости (Value или Brightness)

Яркость	0	0,25	0,50	0,75	1,00
Точность	99,5	99,6	99,4	99,6	100,0

Изменение яркости практически не влияет на точность классификации.

Табл. 5. Влияние на точность классификации сдвига цветового тона (Hue)

Контрастность	0	0,25	0,50	0,75	1,00
Точность	99,6	99,6	99,7	99,5	100,0

Сдвиг цветового тона не приводит к изменению точности классификации.

Табл. 6. Влияние на точность классификации степени компрессии JPEG

Степень компрессии	0	3	6	9	12
Точность	99,8	99,5	99,8	99,6	100,0

Влияние на точность классификации степени компрессии JPEG практически не достоверно.

Заметное уменьшение точности классификации начинается с размера снимка 20×30, т.е. площади снимка, равной 600 пикселям (рис. 6).

Выводы

Продемонстрировано успешное использование свёрточных нейросетей для оценки коммерческой успешности разножанровых фотографий из подборки одного автора, состоящей всего из 400 изображений для обучения и теста, что не противоречит методам глубокого обучения [23, 25]. Кроме того, описанный метод основан на анализе самих изображений и не привязан к сопутствующей информации из социальных сетей и фотобанков, кроме количества покупок каждого конкретного изображения.

Полученные результаты показывают, что при оптимальном выборе архитектуры, алгоритмов обучения и гиперпараметров системы глубокого обучения показывают высокую эффективность использования для оценки степени коммерческой значимости фотографий вплоть до уменьшенных изображений размером 20×30 пикселей с точностью не хуже 98 %.

Продемонстрированный алгоритм будет полезен при разработке маркетинговых стратегий и технологий продаж, и в особенности для облегчения труда фотографов по обработке отснятого материала. В силу специфики архитектуры сети и сформулированной задачи, он показывает большую результативность, нежели некоторые коммерческие проекты в открытом доступе [19, 20], что

говорит о необходимости различать подходы для оценки коммерческой и социальной популярности.



Рис. 6. Зависимость точности классификации от площади изображения

Высокая точность классификации, по всей видимости, обусловлена репрезентативностью выборок и качеством признаков, которыми эти образцы обладают.

Благодаря применению обученной на описанной выше выборке нейросети, вновь размещённые на фотостоках фотографии за 8 месяцев имеют показатели продаж до 20 % выше среднего по предыдущей подборке.

При минимальной доработке данную систему можно использовать для решения других задач, применяя для дообучения соответствующие наборы графических данных.

Благодарности

Работа выполнена при поддержке программы «Повышение конкурентоспособности ведущих университетов РФ» (проект 5-100), контракт №02.a03.21.0005, 27.08.2013

Литература

1. **Aloufi, S.** On the prediction of Flickr image popularity by analyzing heterogeneous social sensory data / S. Aloufi, S. Zhu, A. El Saddik // Sensors. – 2017. – Vol. 17. – 631.
2. **Ellett, J.** New AI-based tools are transforming social media marketing [Электронный ресурс] – 2017. – URL: <https://www.forbes.com/sites/johnellett/2017/07/27/new-ai-based-tools-are-transforming-social-media-marketing/#7437b17669a2>.
3. **More, V.** Study on aesthetic analysis of photographic images techniques to produce high dynamic range images / V. More, P. Agrawal // International Journal of Computer Applications. – 2017. – Vol. 159, No 8. – P. 34-38.
4. **Talebi, H.** NIMA: Neural image assessment / H. Talebi, P. Milanfar // IEEE Transactions on Image Processing. – 2018. – Vol. 27, Issue 8. – P. 3998-4011.
5. **Lu, X.** Deep multi-patch aggregation network for image style, aesthetics, and quality estimation / X. Lu, Z. Lin, X. Shen, R. Mech, J.Z. Wang // Proceedings of the IEEE International Conference on Computer Vision. – 2015. – P. 990-998.
6. **Kang, L.** Convolutional neural networks for no-reference image quality assessment / L. Kang, P. Ye, Y. Li, D. Doermann // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2014. – P. 1733-1740.
7. **Xue, W.** Learning without human scores for blind image quality assessment / W. Xue, L. Zhang, X. Mou // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2013. – P. 995-1002.

8. **Никонов, А.В.** Реконструкция изображений в дифракционно-оптических системах на основе свёрточных нейронных сетей и обратной свёртки / А.В. Никонов, М.В. Петров, С.А. Бибииков, В.В. Кутикова, А.А. Морозов, Н.Л. Казанский // Компьютерная оптика. – 2017. – Т. 41, № 6. – С. 875-887. – DOI: 10.18287/2412-6179-2017-41-6-875-887.
9. **Rubio, F.** Drawing a baseline in aesthetic quality assessment / F. Rubio, M.J. Flores, J.M. Puerts // Proceedings of SPIE. – 2017. – Vol. 10696. – 106961M.
10. **Li, Y.** Image aesthetic quality evaluation using convolution neural network embedded learning / Y. Li [et al.] // Optoelectronics Letters. – 2017. – Vol. 13, Issue 6. – P. 471-475.
11. **Murray, N.** AVA: A large-scale database for aesthetic visual analysis / N. Murray, L. Marchesotti, F. Perronnin // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2012. – P. 2408-2415.
12. **Marchesotti, L.** Assessing the aesthetic quality of photographs using generic image descriptors / L. Marchesotti, F. Perronnin, D. Larlus, G. Csuska // 2011 International Conference on Computer Vision. – 2011. – P. 1784-1791.
13. **Рыцарев, И.А.** Кластеризация медиаконтента из социальных сетей с использованием технологии BigData / И.А. Рыцарев, Д.В. Кирш, А.В. Куприянов // Компьютерная оптика. – 2018. – Т. 42, № 5. – С. 921-927. – DOI: 10.18287/2412-6179-2018-42-5-921-927.
14. **Luo, W.** Content-based photo quality assessment / W. Luo, X. Wang, X. Tang // 2011 International Conference on Computer Vision. – 2011. – P. 2206-2213. – DOI: 10.1109/ICCV.2011.6126498.
15. **Luo, Y.** Photo and video quality evaluation: Focusing on the subject / Y. Luo, X. Tang // European Conference on Computer Vision. – 2008. – P. 386-399. – DOI: 10.1007/978-3-540-88690-7_29.
16. **Sharda, R.** Predicting box-office success of motion pictures with neural networks / R. Sharda, D. Delen // Expert Systems with Applications. – 2006. – Vol. 30. – P. 243-254.
17. **Ning, X.** Rating prediction via generative convolutional neural networks-based regression / X. Ning [et al.] // Pattern Recognition Letters. – 2018. – In Press.
18. **Loureiro, A.L.D.** Exploring the use of deep neural networks for sales forecasting in fashion retail / A.L.D. Loureiro, V.L. Miguéis, L.F.M. da Silva // Decision Support Systems. – 2018. – Vol. 114. – P. 81-93.
19. The parallel dots image recognition service [Electronical Resource]. – URL: <https://www.paralldots.com/object-recognizer> (request date 5.09.2019).
20. EveryPixel aesthetics service [Electronical Resource]. – URL: <https://www.everypixel.com/aesthetics> (request date 5.09.2019).
21. **Kim, P.** MATLAB deep learning: With machine learning, neural networks and artificial intelligence / P. Kim. – Apress, 2017.
22. **Гудфеллоу, Я.** Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль. – пер. с англ. – 2-е изд. – М.: ДМК Пресс, 2018. – 652 с.
23. **Рашка, С.** Python и машинное обучение / С. Рашка. – пер. с англ. – М.: ДМК Пресс, 2017. – 418 с.
24. **Коэльо, Л.П.** Построение систем машинного обучения на языке Python / Л.П. Коэльо, В. Ричарт. – 2-е изд. – пер. с англ. – М.: ДМК Пресс, 2016. – 302 с.
25. **Шолле, Ф.** Глубокое обучение на Python / Ф. Шолле. – СПб: Питер, 2018. – 400 с.

Сведения об авторах

Ефремцев Вадим Григорьевич, 1950 года рождения, в 1972 году окончил Коломенский педагогический институт по специальности «Физика», к.ф.-м.н. (1981). Область научных интересов: компьютерная обработка изображений, программирование, нейронные сети. E-mail: evadim07@gmail.com.

Ефремцев Николай Григорьевич, 1950 года рождения, в 1972 году окончил Коломенский педагогический институт по специальности «Физика». Область научных интересов: компьютерная обработка изображений, программирование, нейронные сети. E-mail: enick07@gmail.com.

Тетерин Евгений Петрович, 1948 года рождения, в 1974 году окончил Всесоюзный заочный машиностроительный институт по специальности «Приборы точной механики», зав. кафедрой физики Ковровской государственной технологической академии им. В.А.Дегтярева, д.т.н. (2004 г.). Область научных интересов: электромагнитное поле, в том числе фоновое, во взаимодействии с веществом, физика жидкого состояния, нейронные сети. E-mail: phys-tep@mail.ru.

Тетерин Петр Евгеньевич, 1987 года рождения, в 2010 году окончил Национальный исследовательский ядерный университет «МИФИ» по специальности «Физика конденсированного состояния вещества», к.ф.-м.н. (2013), работает ведущим инженером в НИЯУ МИФИ. Область научных интересов: детекторы элементарных частиц, осаждение тонких плёнок, нейронные сети. E-mail: peteterin@mephi.ru.

Гансовский Владислав Викторович, 1977 года рождения, в 2000 году окончил Московский энергетический институт по специальности «Инженер электрического транспорта». Работает фотографом-дизайнером. Область научных интересов: профессиональная фотография, компьютерная обработка изображений, нейронные сети. E-mail: vladgans@mail.ru.

ГРНТИ: 28.23.15

Поступила в редакцию 24 января 2019 г. Окончательный вариант – 11 сентября 2019 г.

Deep learning application for box-office evaluation of images

V.G. Efremtsev¹, N.G. Efremtsev¹, E.P. Teterin², P.E. Teterin³, V.V. Gantsovsky¹

¹Independent researcher,

²Kovrov State Technological Academy named after V.A.Degtyarev, Kovrov, Vladimir region, Russia;

³National Research Nuclear University "MEPhI", Moscow, Russia

Abstract

The possibility of application a convolutional neural network to assess the box-office effect of digital images is reviewed. We studied various conditions for sample preparation, optimizer algorithms, the number of pixels in the samples, the size of the training sample, color schemes, compression quality, and other photometric parameters in view of effect on training the neural network. Due to the proposed preliminary data preparation, the optimum of the architecture and hyperparameters of the neural network we achieved a classification accuracy of at least 98%.

Keywords: deep learning, neural networks, image analysis.

Acknowledgements: Authors thank for the support from National Research Nuclear University MEPhI in the framework of the Russian Academic Excellence Project (contract No. 02.a03.21.0005, 27.08.2013).

Citation: Efremtsev VG, Efremtsev NG, Teterin EP, Teterin PE, Gantsovsky VV. Deep learning application for box-office evaluation of images. *Computer Optics* 2020; 44(1): 127-132. DOI: 10.18287/2412-6179-CO-515.

References

- [1] Aloufi S, Zhu S, El Saddik A. On the prediction of Flickr image popularity by analyzing heterogeneous social sensory data. *Sensors* 2017; 17: 631.
 - [2] Ellett J. New AI-based tools are transforming social media marketing 2017. Source: <https://www.forbes.com/sites/johnellett/2017/07/27/new-ai-based-tools-are-transforming-social-media-marketing/#7437b17669a2>.
 - [3] More V, Agrawal P. Study on aesthetic analysis of photographic images techniques to produce high dynamic range images. *Int J Comp Applicat* 2017; 159(8): 34-38.
 - [4] Talebi H, Milanfar P. NIMA: Neural image assessment // *IEEE Trans Image Process* 2018; 27(8): 3998-4011.
 - [5] Lu X, Lin Z, Shen X, Mech R, Wang JZ. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. *Proc IEEE Int Conf Comput Vis* 2015: 990-998.
 - [6] Kang L, Ye P, Li Y, Doermann D. Convolutional neural networks for no-reference image quality assessment, in *Proc IEEE Conf Comput Vis Pattern Recognit* 2014: 1733-1740.
 - [7] Xue W, Zhang L, Mou X. Learning without human scores for blind image quality assessment. *Proc IEEE Conf Comput Vis Pattern Recognit* 2013: 995-1002.
 - [8] Nikonorov AV, Petrov MV, Bibikov SA, Kutikova VV, Morozov AA, Kazanskiy NL. Image restoration in diffractive optical systems using deep learning and deconvolution. *Computer Optics* 2017; 41(6): 875-887. DOI: 10.18287/2412-6179-2017-41-6-875-887.
 - [9] Rubio F, Flores MJ, Puerts JM. Drawing a baseline in aesthetic quality assessment. *Proceedings of SPIE*. – 2017. – Vol. 10696. – 106961M.
 - [10] Li Y, et al. Image aesthetic quality evaluation using convolutional neural network embedded learning. *Optoelectron Lett* 2017; 13(6): 471-475.
 - [11] Murray N, Marchesotti L, Perronnin F. AVA: A large-scale database for aesthetic visual analysis. *Proc IEEE Conf Comput Vis Pattern Recognit* 2012: 2408-2415.
 - [12] Marchesotti L, Perronnin F, Larlus D, Csurka G. Assessing the aesthetic quality of photographs using generic image descriptors. *ICCV* 2011: 1784-1791.
 - [13] Rycarev IA, Kirsh DV, Kupriyanov AV. Clustering of media content from social networks using bigdata technology. *Computer Optics* 2018; 42(5): 921-927. DOI: 10.18287/2412-6179-2018-42-5-921-927.
 - [14] Luo W, Wang X, Tang X. Content-based photo quality assessment. *ICCV* 2011: 2206-2213. DOI: 10.1109/ICCV.2011.6126498.
 - [15] Luo Y, Tang X. Photo and video quality evaluation: Focusing on the subject. *ECCV* 2008: 386-399. DOI: 10.1007/978-3-540-88690-7_29.
 - [16] Sharda R, Delen D. Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications* 2006; 30: 243-254.
 - [17] Ning X, et al. Rating prediction via generative convolutional neural networks based regression. *Pattern Recognit Lett* 2018; In Press.
 - [18] Loureiro ALD, Miguéis VL, da Silva LFM. Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems* 2018; 114: 81-93.
 - [19] The parallel dots image recognition service. Source: <https://www.paralleldots.com/object-recognizer>.
 - [20] EveryPixel aesthetics service. Source: <https://www.everypixel.com/aesthetics>.
 - [21] Kim P. MATLAB deep learning: With machine learning, neural networks and artificial intelligence. Apress; 2017.
 - [22] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press, 2016.
 - [23] S.Raschka Python Machine Learning. Birmingham: Packt Publishing; 2015.
 - [24] Richert W, Coelho LP. Building machine learning systems with Python. Packt Publishing; 2013.
 - [25] Chollet F. Deep learning with Python. New York: Manning Publications; 2017.
-

Authors' information

Vadim Grigorievich Efremtsev, (b.1950) graduated from Kolomna Pedagogical Institute in 1972 with a degree in Physics. Ph.D. (1981). Research interests: computer image processing, programming, neural networks. E-mail: evadim07@gmail.com.

Nikolay Grigorievich Efremtsev, (b.1950) graduated from Kolomna Pedagogical Institute in 1972 with a degree in Physics. Research interests: computer image processing, programming, neural networks. E-mail: enick07@gmail.com.

Evgeniy Petrovich Teterin, (b.1948) graduated from the All-Union Correspondence Mechanical Engineering Institute in 1974 with a degree in Precision Mechanics Devices. Currently he works as head of Physics department in Kovrov State Technological Academy named after V.A. Degtyarev. Dr. of Science (2004). Research interests: electromagnetic fields incl. backgrounds and interaction with matter, liquid state physics, acoustics, neural networks. E-mail: phys-tep@mail.ru.

Petr Evgenyevich Teterin (b. 1987) graduated from National Research Nuclear University "MEPhI" in 2010, Solid State Physics department. Ph.D. (2013). He works as the lead engineer at MEPhI. Research interests are currently focused on elementary particle detectors, thin film deposition and data analysis, neural networks. E-mail peteterin@mephi.ru.

Vladislav Viktorovich Gantsovsky (b.1977) graduated from Moscow Power Engineering Institute in 2000 with a degree in Electrical Transport Engineering. Works as a photographer designer. Research interests: professional photography, computer image processing, neural networks. E-mail: vladgans@mail.ru.

Received January 24, 2019. The final version – September 11, 2019.
