

31. Mann Y., CohenOr D. Selective pixel transmission for navigating in remote virtual environments. URL: <https://www.cs.tau.ac.il/~dcor/articles/older/> (дата обращения 01.03.2018).

32. Fechteler P., Eisert P. Depth map enhanced macroblock partitioning for H. 264 video coding of computer graphics content. URL: <http://iphome.hhi.de/eisert/papers/> (дата обращения 01.03.2018).

33. Moser M., Weiskopf D. Interactive Volume Rendering on Mobile Devices. URL: <https://pdfs.semanticscholar.org/f917/> (дата обращения 01.03.2018).

34. Using expressive rendering for remote visualization of large city models / J. C. Quillet, G. Thomas, X. Granier, P. Guitton, J. E. Marvie // URL: <https://hal.inria.fr/inria-00106832v2/document/> (дата обращения 21.02.2018).

T. V. Gerasimova

Saint Petersburg Electrotechnical University «LETI»

WEB AND 3D GRAPHICS

The increasing role of 3D graphics as an important part of multimedia web content is considered. In the proposed work, a review of the current state of the art of using real-time 3D graphics in the Internet, covering rendering methods, methods for describing a scene and providing 3D data, is presented and discussed in detail. Approaches to 3D-rendering based on the browser, classified by the level of declarative behavior are presented. The evolution of the X3D standard is shown, which allows to obtain a lower level of access to graphic equipment, which constantly increases the power. Considered remote rendering methods allow you to transfer high-quality 3D graphics to a wide range of devices, and in recent years also widely studied methods of delivering content for web-based 3D applications. All these developments are reflected in the increase in the number of areas for which applications for the 3D network are being developed.

3D, browser, plugin, HTML5, WebGL, X3D, X3DOM, XML, Three.js, Java, web, rendering, canvas

УДК 681.3

А. М. Голубков

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Бинарная классификация изображений на примере задачи распознавания лиц

Рассматривается практическое применение теоремы о разделимости, утверждающей, что в многомерных пространствах произвольно выбранная точка этого пространства отделима линейным дискриминантом от любого произвольно выбранного множества точек этого пространства при достаточно большой размерности. Показано, что «достаточно большой» является размерность больше двадцати. В качестве приложения данной теоремы приводится задача распознавания (бинарной классификации) лиц людей. Используя сверточные нейронные сети, для каждой фотографии лица формируется вектор признаков (в данном случае 128-мерный). Каждый такой вектор можно рассматривать как точку в 128-мерном пространстве. Показано, что, используя теорему о разделимости, можно с вероятностью, близкой к единице, отделить лицо одного человека от лиц остальных людей. В качестве дискриминанта используется линейный дискриминант Фишера, который предоставляет одношаговую процедуру отделения, не требующую предварительного обучения. В результате экспериментов на наборе из 13 233 фотографий была получена точность 0.9980 с полнотой 0,9623.

Нейронные сети, бинарная классификация, теорема о разделимости, линейный дискриминант, расстояние Махаланобиса

Бинарная классификация изображений является популярной задачей со множеством технических приложений. Решение данной задачи отве-

чает на вопрос, относится ли данное изображение к заданному классу или нет. Например, для актуальной задачи автопилотирования автомобилей

задача бинарной классификации может ответить на вопрос, находится ли на изображении камеры автомобиля человек или нет, т. е. относится ли изображение к классу 1 (человек присутствует) или к классу 0 (человек отсутствует).

Существует множество способов решения данной задачи, начиная от классических методов машинного зрения и заканчивая современными методами машинного обучения и нейронных сетей. В частности, новейшие архитектуры нейронных сетей показывают точность, превосходящую человеческую на множестве задач классификации (не только бинарной). С другой стороны, преобладающий на сегодняшний день сверточный тип нейронных сетей требует большого массива данных (изображений) как класса 1, так и класса 0 для обучения сети с целью получения необходимого уровня точности.

Ниже представлен метод, основанный на теореме о разделимости, опубликованной в августе 2017 г., а также на свойствах многомерных пространств и особенностях работы сверточных нейронных сетей.

Архитектура и принцип работы сверточных нейронных сетей. Сверточные нейронные сети [1] (далее СНС) являются типом нейронных сетей, показывающим наилучшие результаты при работе с изображениями (в отличие от, например, рекуррентных нейронных сетей, работающих преимущественно с последовательными сигналами, например звуком). Общая архитектура СНС представлена на рис. 1.

На вход нейронной сети поступает изображение, представленное в виде слоя значений интенсивности каждого пикселя изображения (если изображение черно-белое) или в виде трех слоев, соответствующих каждому цвету в схеме RGB (каждый слой – интенсивность пикселя в соответствующем цвете). Далее полученные значения проходят множество нелинейных преобразований (за которые отвечают нелинейные активационные функции нейронов), после чего поступают на по-

следний нейрон, результатом работы которого является вероятность принадлежности изображения к классу 1.

Для правильной работы последнего нейрона (оценки вероятности принадлежности к классу 1) СНС необходимо обучить на множестве размеченных данных, т. е. таких, где каждому изображению заранее проставлен соответствующий класс. Такой набор данных называется тренировочным датасетом. Чем большая точность ответа необходима, тем большим должен быть размер датасета, причем распределение классов изображений в нем должно быть примерно одинаковым (т. е. ~50 % изображений класса 1 и ~50 % изображений класса 0). Это может представлять большую сложность в практических задачах, например в задаче распознавания лиц, когда необходимо определить, является ли данный человек человеком, фотография которого имеется в базе данных, или нет. В базе данных может быть огромное количество фотографий других людей (класс 0), но при этом только несколько (чаще всего 1–3) фотографий конкретного человека (класс 1). Более того, каждый раз придется распознавать разных людей, т. е. каждый раз переобучать нейронную сеть на распознавание конкретного человека, что очень долго (или иметь набор заранее обученных нейронных сетей для каждого человека, что потребует очень много памяти).

Таким образом, необходимо решить задачу быстрого и не требующего много памяти распознавания конкретного человека, используя при этом способность СНС к анализу изображений.

Признаковое описание изображений. Назовем *состоянием сети* значения выходов всех нейронов на произвольно выбранном слое. Вектор, составленный из этих значений, является *вектором состояния*. Исходя из особенностей работы СНС он описывает изображение, которое нейронная сеть обрабатывает в данный момент. Каждый элемент в векторе состояния является определенным признаком изображения. Таким

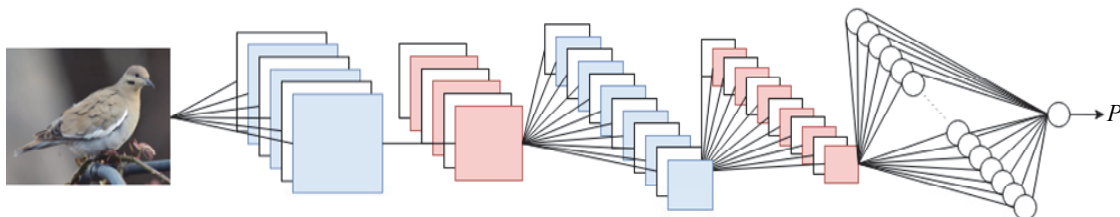


Рис. 1

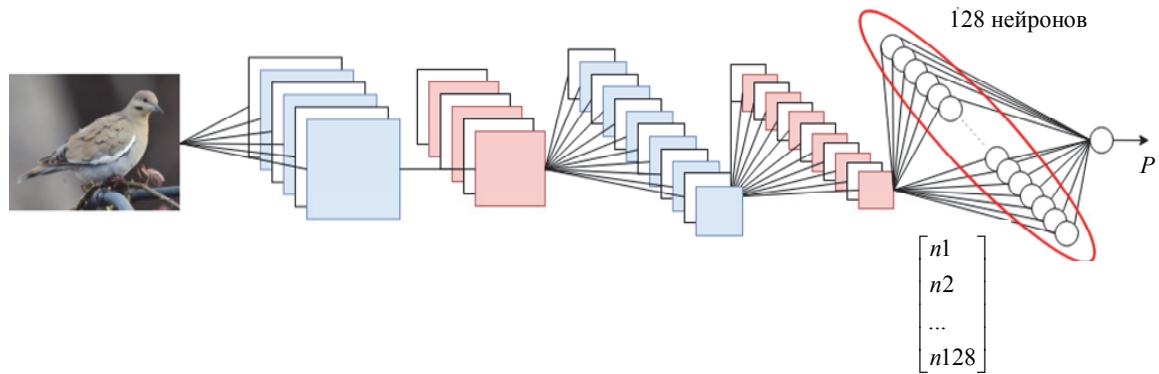


Рис. 2

образом, данный вектор называется также *вектором признаков*.

Идея признакового описания изображения состоит в том, чтобы «снять» вектор состояния с предпоследнего слоя нейронной сети, т. е. со слоя, предшествующего слою с одним нейроном (рис. 2), для каждого изображения.

В ходе своей работы СНС распознает элементы изображения, двигаясь от простых элементов (прямые и изогнутые линии, точки) на первых слоях и заканчивая более сложными, составными элементами на последних слоях (применительно к лицу человека это глаза, нос, уши и другие сложные формы). Таким образом, формируя вектор признаков с помощью самого глубокого слоя (предпоследний слой), получается вектор признаков, описывающий максимально абстрактные элементы изображения. С помощью данного метода формируется числовое (признаковое) описание изображения, используя преимущества сверточных нейронных сетей.

Отделение точек от множества в многомерных пространствах. Теорема о разделимости утверждает, что в пространствах с достаточно большой размерностью любая точка может быть отделима линейным функционалом от остального множества точек с вероятностью [2]:

$$P(M, n) \geq \max_{\varepsilon \in (0,1)} (1 - (1 - \varepsilon)^n) \left(1 - \frac{\rho(\varepsilon)^n}{2} \right)^M, \quad (1)$$

где $\rho(\varepsilon) = (1 - (1 - \varepsilon)^2)^{1/2}$; ε – регулируемый параметр, подбираемый экспериментально к каждому распределению, для распределения на сфере обычно равен $\frac{1}{5}$ [2]; M – множество одинаково

распределенных точек; M – мощность данного множества; n – размерность пространства.

Как видно из (1), вероятность отделения точки от множества растет экспоненциально с увеличением размерности пространства. Например, при $n = 25$ вероятность отделения точки $\left(\varepsilon = \frac{1}{5} \right)$:

$$P(M, 25) \geq \left[1 - \left(1 - \frac{1}{5} \right)^n \right] \left(1 - \frac{0.6^n}{2} \right)^M = 0.988470 \cdot 0.999981^M.$$

Как видно из неравенства, уже при $n = 25$ вероятность отделения точки близка к единице. При этом вероятность растет экспоненциально с увеличением n (рис. 3).

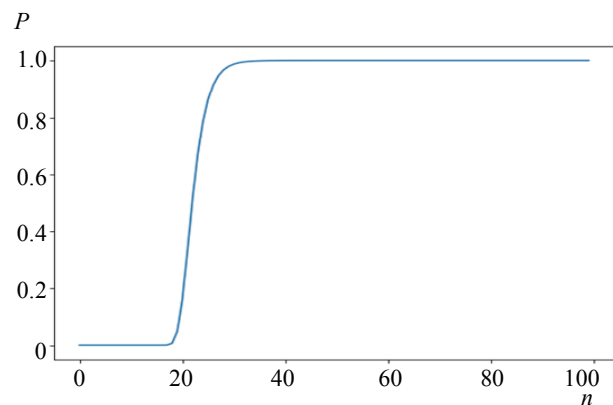


Рис. 3

Отделимость точки осуществляется линейным функционалом. В качестве такого функционала можно использовать линейный дискриминант Фишера или однонейронную сеть (или любой другой линейный дискриминант).

Используя линейный дискриминант Фишера, получаем следующее неравенство [2]:

$$l_y(x) = \left\langle \frac{\mathbf{w}(y, M)}{\|\mathbf{w}(y, M)\|}, x \right\rangle - c < 0, \quad (2)$$

$$\mathbf{w}(y, M) = \Sigma^{-1} \left(y - \frac{1}{M-1} \sum_{x \neq y} x \right),$$

где $\mathbf{w}$ – вектор коэффициентов дискриминанта Фишера; y – отделяемый элемент; x – любой элемент из множества M ; Σ – ковариационная матрица множества M ; c – параметр, задающий границу «разделимости», т. е. при каком минимальном c элементы можно считать разделимыми.

Распознавание человеческих лиц по фотографии. Теорема о разделимости утверждает, что при достаточно больших размерностях пространства любая точка этого пространства с вероятностью, близкой к единице, отделима от остального множества. Таким образом, если представить, что векторное описание изображения с количеством признаков n и есть точка в пространстве размерностью n , то используя линейный дискриминант Фишера (или любой другой линейный дискриминант) возможно отделение данной точки (признаковое описание фотографии лица) от всего остального множества (других признаков описаний фотографий лиц). При этом нейронную сеть не нужно обучать на конкретных необходимых лицах, достаточно получить признаковое описание этого лица.

Существует множество зарекомендовавших себя архитектур СНС. Для проведения экспериментов была выбрана архитектура Inception ResNet v1 [3], разработанная компанией Google, заранее предобученная на большом наборе неизвестных лиц.

Для проведения экспериментов была использована база фотографий человеческих лиц LFW (Labeled Faces in the Wild), в которой собрано 13 233 фотографии знаменитых людей. Несмотря на то, что существует множество специально сформированных и предварительно обработанных баз фотографий лиц, был выбран набор фотографий людей «in the wild», т. е. с лицами под

разными углами, под разным освещением и с разным фоном. Это было необходимо для оценки работы метода на реальных задачах.

Чтобы распознать нужного человека, необходимо совершить следующие шаги:

1. Предобработка базы LFW. Так как фотографии лиц сделаны «in the wild», т. е. в живой среде, необходимо каждую фотографию обрезать таким образом, чтобы на ней осталось только лицо человека. При этом получившиеся изображения должны быть квадратными и одного и того же размера (в данном случае размером 128×128 пикселей). Для этой цели используется отдельная нейронная сеть.

2. После предобработки каждое изображение подается на вход сверточной нейронной сети Inception ResNet v1, при этом работа сети останавливается не на последнем слое, а на предпоследнем (рис. 2), получая тем самым признаковое описание каждого изображения размеров 128 признаков.

3. Выбирается нужный человек из базы данных, выделяется одна его фотография (признаковый вектор), после чего строится дискриминант Фишера по формуле (2) с *любой другой* фотографией, при этом выбирается оптимальный порог c (наилучшим значением порога в данном эксперименте оказалось $c = 0.7$). Если «любая другая» фотография принадлежит тому же человеку, для которого строился дискриминант, то его значение будет больше нуля ($l_y(x) > 0$), иначе меньше.

Разработка и тестирование метода проводилась в среде Anaconda на языке программирования Python с использованием библиотеки машинного обучения TensorFlow.

Таким образом, модель системы обработки фотографии выглядит следующим образом (рис. 4).

Точность распознавания лиц описанным методом на датасете LFW приведена в таблице (FNR – False Negative Rate, частота появления

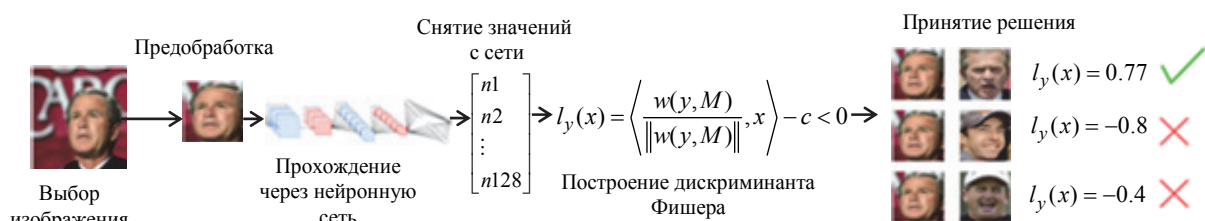


Рис. 4

ошибок ложного срабатывания, т. е. если две фотографии принадлежат одному и тому же человеку, но система «сказала», что это разные люди; FPR – False Positive Rate, частота появления ошибок ложного срабатывания, т. е. когда фотографии принадлежат разным людям, но система «сказала», что это один и тот же человек).

Количество векторов	Средний FNR	Средний FPR	Средняя точность	Средняя полнота
13 233	0.0377	0.0078	0.9980	0.9623

В качестве практического применения теоремы о разделимости была рассмотрена задача распознавания человеческих лиц. Было показано, что для качественного распознавания достаточно лишь правильно подобрать порог c (в эксперименте он был равен 0.7), а также была необходима лишь одна фотография распознаваемого лица.

В качестве дальнейших исследований рассматриваются следующие направления:

1. Работа с множеством, разделенным на кластеры, так как это может усложнить работу линейного дискриминанта.

2. Применение различных видов линейных дискриминантов, в том числе одиночного нейрона.

3. Дополнительная предобработка векторов признаков для возможного ускорения расчета линейного дискриминанта.

Теорема о разделимости позволяет решать широкий класс задач, использующих многомерное описание объектов. Нейронные сети, которые оперируют многомерными векторами, отлично подходят под эти условия, что дает возможность дальнейшего изучения и развития практического применения свойства отделимости точек в многомерных пространствах.

СПИСОК ЛИТЕРАТУРЫ

1. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet Classification with Deep Convolutional Neural Networks. URL: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (дата доступа 21.03.18).

2. One-trial correction of legacy AI systems and stochastic separation theorems / A. N. Gorban, I. Romanenko, R. Burton, I. Y. Tyukin. URL: <https://arxiv.org/abs/1610.00494> (дата доступа 12.03.18).

3. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning / C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi. URL: <https://arxiv.org/pdf/1602.07261> (дата доступа 21.03.18).

A. M. Golubkov

Saint Petersburg Electrotechnical University «LETI»

FACE RECOGNITION USING IMAGES BINARY CLASSIFICATION METHODS.

Stochastic separation theorem applied to face recognition task using Fisher's linear discriminant is presented in this paper. The theorem says that in spaces with high dimension there is a high point dividing probability. There is an idea about using feature vectors as a points in a high dimension spaces. As a method to take feature vectors convolutional neural networks are used. In this paper Inception ResNet v1 neural network architecture is used with 128-dimension feature vectors. As it can be seeing, precision of this method is equal to 0.9980 and recall is equal to 0.9623.

Neural networks, binary classification, stochastic separation theorem, linear discriminant, Mahalanobis distance