УДК 004.93'12

А.А. Хуршудов, В.Н. Марков

СПОНТАННОЕ ВЫДЕЛЕНИЕ ИЕРАРХИИ ДВУМЕРНЫХ ПРИЗНАКОВ ДЛЯ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ

A.A. Khurshudov, V.N. Markov

UNSUPERVISED LEARNING OF HIERARCHICAL 2D FEATURES FOR IMAGE CLASSIFICATION

В задачах классификации и распознавания изображений одной из ключевых проблем является выделение компонентов (признаков), определяющих категорию классификации, устойчивых к инвариантным преобразованиям изображенного объекта. Предложен эффективный способ построения расширяемой многоуровневой модели, способной выделять такие признаки из обучающей выборки без использования учителя.

Рассмотрены преимущества иерархического подхода к выделению признаков и его способность к инкапсулированию структурно сложных компонентов изображения, обработка которых представляет значительные вычислительные трудности с использованием классических обучающих алгоритмов. Метод оперирует на плоских (двумерных) изображениях, но обладает потенциальной возможностью расширения для работы с трехмерными объектами. Полученная модель может использоваться в качестве детектора признаков для множества различных методов обучения, таких как сверточные нейронные сети.

ИЕРАРХИЧЕСКАЯ МОДЕЛЬ; ГЛУБОКОЕ ОБУЧЕНИЕ; СПОНТАННОЕ ВЫДЕЛЕНИЕ ПРИЗНАКОВ; ОБНАРУЖЕНИЕ ПРИЗНАКОВ.

One of the key problems of image classification and pattern recognition domains is that of feature detection. The desired features are expected to be robust and invariant to a number of spatial transformations, compact enough to evade the «curse of dimensionality» which is a frequent obstacle when dealing with large natural images, and provide a characteristic relation to a classification category with high probability. There exists a number of approaches developed to reach the stated goals, including a variety of deep learning models, such as Restricted Boltzman Machines, convolutional networks, autoencoders, PCA, Deep Belief Networks, etc. However, most applications of the above-mentioned algorithms are often concentrated on obtaining the most accurate features for a chosen dataset rather than trying to extract the inner structure of the data. This paper suggest a slightly different approach, namely a method for building a hierarchy of meaningful features with each level composed of the features from a previous layer. Such model has multiple applications — it can serve as a composite feature detector in an unsupervised pre-training step of learning, or be itself a metric that answers the question of whether the same spatial structure is present across the dataset. The proposed approach exploits the idea of local connectivity supposing that multiple adjacent image parts which contain some meaningful features might present another, more high-level feature when composed together. We also discuss the advantages of a hierarchical feature model, such as the ability to guess a high-level feature presence by discovering a collection of low-level features concentrated in the same area, or its stability against noise and distortion which happens due to the fact that each feature level accepts a certain degree of deviation accumulating those to the top of the hierarchy. The resulting model operates

on 2D images, but can be easily extended in order to extract 3D features from a continuous data input, such as a movie, which promises to be a good way to deal with 3D transformations, which can drastically change the appearance of an object while preserving its identity.

HIERARCHICAL MODEL; DEEP LEARNING; UNSUPERVISED FEATURE LEARNING; FEATURE DETECTION.

Задача классификации объектов, представленных в виде изображений, состоит из конечного набора решений, каждое из которых соотносит выбранное изображение с соответствующей категорией классификации. Общепринятым подходом для принятия таких решений является метод выделения признаков, который используют как системы компьютерного зрения, так и естественный интеллект животных и человека [1]. Метод выделения признаков в свою очередь состоит в допущении того, что каждой категории классификации соответствуют некоторые устойчивые, повторяющиеся характеристики, которые с высокой вероятностью встречаются в изображениях, принадлежащих данной категории, и с низкой вероятностью – в изображениях других категорий. Однако по настоящее время не существует однозначного формализованного способа нахождения таких характеристик для произвольного набора категорий и изображений. Отдельную проблему представляет собой классификация изображений для задач компьютерного зрения в естественном окружении, с участием трехмерных объектов, эффектов освещения и пространственных преобразований, таких как вращение, масштабирование и трансляция. С учетом эффекта этих преобразований изображения одного и того же объекта могут значительным образом различаться цветом, формой или контурами, что существенно затрудняет распознавание. Нахождение метода выделения признаков, устойчивых к подобным преобразованиям, является глобальной задачей в области компьютерного зрения и распознавания изображений.

Иерархия признаков

Существует некоторое количество аргументов в пользу того, что эффективным кандидатом для искомого метода может быть метод построения многоуровневых иерархий признаков, где более простые элементы изображения, такие как небольшие участки, содержащие штрихи и границы, локально (топографически) объединяются в устойчивые признаки более высокого уровня, представляющие собой контуры, геометрические фигуры и более сложные структурные компоненты. Среди свидетельств, подкрепляющих это предположение, следующие:

- в методах «глубокого обучения» («deep learning») – интенсивно развивающейся современной ветви машинного обучения – утверждается, что обучающиеся модели с преобладанием глубокой структуры, такие как многослойные нейронные сети, обладают большим потенциалом к выражению сложных, структурных признаков, которые неспособна представить одноуровневая модель [2]. Существует значительное количество алгоритмов, использующих этот подход в обучении, таких как ограниченная машина Больцмана, глубокие автоэнкодеры, сверточные нейронные сети, которые показали свою эффективность по сравнению с классическими перцептронами;
- существуют доводы в пользу того, что мозг человека и животных использует построение иерархии признаков. Так, во время классического эксперимента Хьюбела и Визеля по поиску детекторов признаков в зрительной коре головного мозга [3] были обнаружены клетки, реагирующие на определенные сочетания клеток-детекторов признаков более низшего уровня. Организаторы эксперимента сделали предположение о существовании детекторов более высокого уровня и формировании иерархии зрительных признаков.

Для построения такой иерархии необходимо решить две подзадачи:

для отдельно взятого уровня иерархии, начиная с первого, выделить признаки одинаковой структурной сложности;

сформулировать правило группировки признаков и распространения их в более высокие уровни иерархии.

Выделение признаков на отдельном уровне иерархии

Рассмотрим в качестве обучающей выборки случайным образом выбранные фрагменты изображения размера $n \times n$. Представим каждый фрагмент в виде матрицы соответствующего размера, где числа матрицы будут соответствовать интенсивности пикселей оригинального фрагмента. Предположим, что каждую такую матрицу можно представить в виде линейной суммы компонентов следующим образом:

$$x = \sum_{i=1}^{c} a_i x_i, \qquad (1)$$

где x_i – i-я матрица размера $n \times n$; a_i – i-й коэффициент; c – количество компонентов линейной суммы, в общем случае бесконечно большое.

Добавим к этому разложению следующее условие: должно существовать минимальное конечное число коэффициентов a_i, отличных от нуля. Задача нахождения такой суммы для данной матрицы представляет собой задачу разреженной аппроксимации, которая может решаться различными способами, такими как метод покоординатного спуска или метод наименьших углов (LARS) [4].

Полученные в результате матрицы $x_0, x_1, ..., x_c$ будут представлять собой минимальный набор функционально различных (в силу условия разреженности) компонентов, которые могут использоваться для представления любого фрагмента размера $n \times n$. Поиск такого разложения производится на выборке фрагментов оригинального изображения и с ограничением количества компонентов c некоторым эмпирически выбранным значением. Результат разложение показан на рис. 1.

При обработке небольших фрагментов изображения в результате получается набор детекторов границ, которые можно использовать в качестве примитивных признаков контура объекта. Сами по себе такие признаки малоинформативны и не несут информации, существенной для распознавания сложных объектов, однако могут послужить основой для признаков более высокого порядка [6].

Выделение признаков высокого уровня

Полученный набор компонентов позволяет представить любой фрагмент изображения размера $n \times n$ в виде вектора длиной c. Определим функции, позволяющие осуществлять преобразование фрагмента изображения в его кодированное представление, записанное в двумерном виде (в форме матрицы):

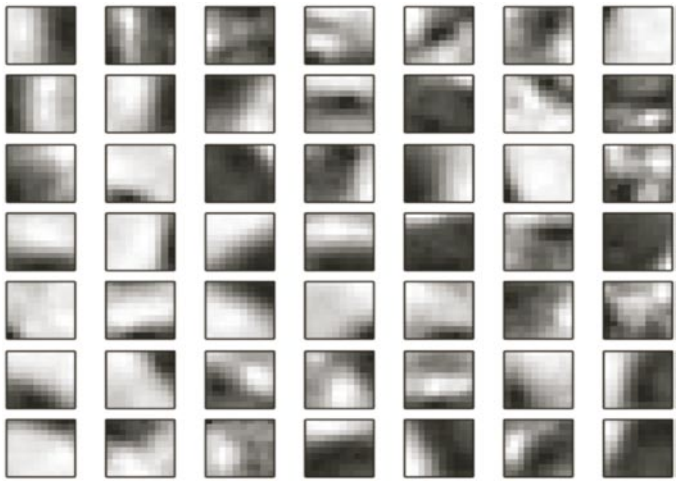

Рис. 1. Компоненты, найденные при обработке выборки фрагментов размера 10×10 (получены с использованием программного пакета scikit-learn [5])

$$Encode\left(\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & & & x_{2n} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}\right) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1\sqrt{c}} \\ a_{21} & & & a_{2\sqrt{c}} \\ \vdots & & & \vdots \\ a_{\sqrt{c}1} & a_{\sqrt{c}2} & \cdots & a_{\sqrt{c}\sqrt{c}} \end{pmatrix}, \quad (2)$$

а также обратное преобразование:

$$Decode\left(\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1\sqrt{c}} \\ a_{21} & & & a_{2\sqrt{c}} \\ \vdots & & & \vdots \\ a_{\sqrt{c}1} & a_{\sqrt{c}2} & \cdots & a_{\sqrt{c}\sqrt{c}} \end{pmatrix}\right) = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & & & x_{2n} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}. \quad (3)$$

Полученные преобразования в общем случае не являются взаимно обратными. Функция *Encode* выполняется с помощью одного из семейств алгоритмов (Orthogonal matching pursuit, LASSO-LARS) [4], задачей которых является поиск наиболее точных

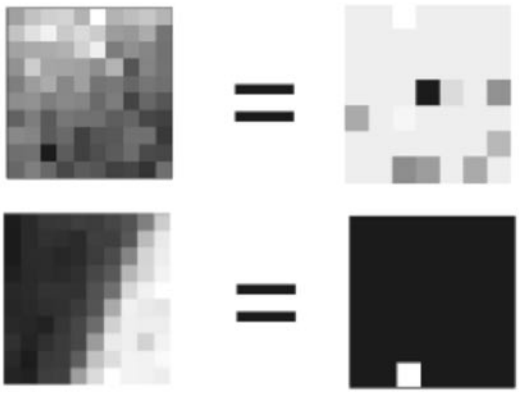

Рис. 2. Визуальное представление преобразования *Encode*
Каждый пиксель правой части представляет собой значение коэффициента a_i при соответствующем компоненте

значений коэффициентов a_i, чтобы линейная сумма соответствующих компонентов наиболее точно (с минимальным отклонением) представляла фрагмент x_{nxn}. Функция *Decode* тождественна формуле (1), с учетом того, что вектор $(a_1, a_2, ..., a_c)$ здесь представлен в форме матрицы размера $\sqrt{c}x\sqrt{c}$. Для обучения второго уровня иерархии выберем случайное количество фрагментов изображения размера m, где $m = kn, k \subset \mathbb{Z}$, после чего к каждому из них применим следующее преобразование:

$$E\left(\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & & & x_{2m} \\ \vdots & & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mm} \end{pmatrix}\right) = \begin{pmatrix} Encode\left(\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix}\right) & \cdots & Encode\left(\begin{pmatrix} x_{1(k-1)n} & \cdots & x_{1kn} \\ \vdots & & \vdots \\ x_{n(k-1)n} & \cdots & x_{nkn} \end{pmatrix}\right) \\ \vdots & & \vdots \\ Encode\left(\begin{pmatrix} x_{(k-1)n1} & \cdots & x_{(k-1)nn} \\ \vdots & & \vdots \\ x_{kn1} & \cdots & x_{knn} \end{pmatrix}\right) & \cdots & Encode\left(\begin{pmatrix} x_{(k-1)n(k-1)n} & \cdots & x_{(k-1)nkn} \\ \vdots & & \vdots \\ x_{kn(k-1)n} & \cdots & x_{knkn} \end{pmatrix}\right) \end{pmatrix}. \quad (4)$$

В ходе этого преобразования фрагмент x_{mxm} разбивается на k^2 малых фрагментов, каждый из которых затем представляется в кодированной форме. Визуализация процесса показана на рис. 3.

Результат преобразования составляет обучающую выборку для второго уровня иерархии признаков, после чего для полученной выборки решается соответствующая задача подбора компонентов (1). Закономер-

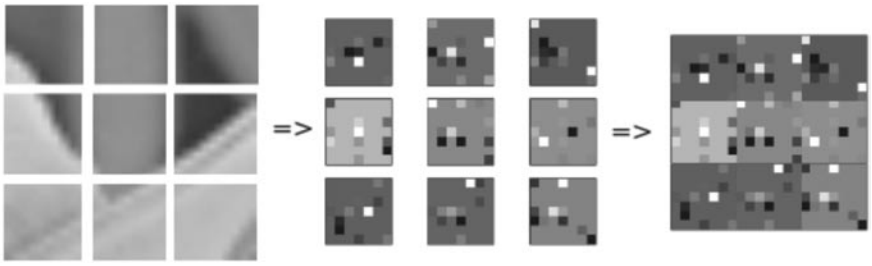

Рис. 3. Визуализация преобразования (4). Фрагмент изображения разбивается на малые фрагменты и кодируется по частям

ности взаимного расположения малых фрагментов будут сохраняться в случае перевода их в кодированное представление и, таким образом, среди преобразованных фрагментов второго уровня будут наблюдаться повторяющиеся структурные элементы, такие как сочетания признаков первого уровня: линии, углы, детали геометрических фигур. Соответствующий принцип позволяет наращивать уровни иерархии и получать признаки более высоких уровней. Тестирование метода демонстрирует способность к отысканию структурно сложных признаков, таких как контуры человеческого лица, за приемлемое время (рис. 4).

Пример демонстрирует успешное соблюдение принципа подбора эффективного числа компонентов. Сохраняется максимальная независимость признаков друг от друга: присутствуют разные формы лиц, цвет кожи, ориентация в пространстве. Любую фотографию обучающей выборки становится возможным представить в виде композиции полученного набора признаков.

Рассмотренный метод имеет следующие достоинства:

вычислительная простота: на аналогичном наборе данных извлечение признаков с помощью иерархии локальных фрагментов изображения выполняется в среднем быстрее, чем с использованием вычислительно дорогостоящих операций, характерных для сверточных сетей и RBM;

масштабируемость: благодаря обработке локальных участков изображения метод может использоваться на изображениях любого размера. Единственное условие – возможность охватить за конечное число уровней иерархии фрагменты, содержащие искомые структурные признаки;

способность обучаться без учителя на

Рис. 4. Признаки третьего уровня иерархии, извлеченные из фотографий лиц

однородной выборке (изображений объектов одного типа). Первые уровни иерархии способны извлекать признаки даже из случайной выборки, концентрируясь на элементах геометрических форм;

эксплуатация принципов глубокого обучения – извлеченные иерархией признаки соответствуют отдельным структурным составляющим объекта. Такие признаки могут использоваться для выделения частей в сложных, составных объектах, таких как элементы лица.

Недостаток метода на данном этапе – необходимость ручной настройки метапараметров, отвечающих за размер фрагмента и количество выделяемых компонентов.

Перспективным направлением развития является обучение на выборке различных проекций трехмерных объектов для формирования признаков, инвариантных к пространственным преобразованиям в трехмерном пространстве.

СПИСОК ЛИТЕРАТУРЫ

1. **Morrone M.C., Burr D.C.** Feature detection in human vision: A phase-dependent energy model // Proc. of the Royal Society of London. Ser. B. Biological sciences. 1988. Pp. 221–245.

2. **Bengio Y.** Learning deep architectures for AI //Foundations and trends® in Machine Learning. 2009. T. 2. No. 1. Pp. 1–127.

3. **Hubel D.H., Wiesel T.N.** Receptive fields, binocular interaction and functional architecture in the cat's visual cortex // The Journal of physiology. 1962. Vol. 160. No. 1. P. 106.

4. **Kreutz-Delgado K. et al.** Dictionary learning algorithms for sparse representation // Neural computation. 2003. Vol. 15. No. 2. Pp. 349–396.

5. **Pedregosa F. et al.** Scikit-learn: Machine learning in Python // The Journal of Machine Learning Research. 2011. Vol. 12. Pp. 2825–2830.

6. **Erhan D. et al.** Why does unsupervised pretraining help deep learning? // The Journal of Machine Learning Research. 2010. Vol. 11. Pp. 625–660.

REFERENCES

1. **Morrone M.C., Burr D.C.** Feature detection in human vision: A phase-dependent energy model, *Proceedings of the Royal Society of London. Series B, Biological sciences*, 1988, Pp. 221–245.

2. **Bengio Y.** Learning deep architectures for AI, *Foundations and trends® in Machine Learning*, 2009, Vol. 2, No. 1, Pp. 1–127.

3. **Hubel D.H., Wiesel T.N.** Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *The Journal of physiology*, 1962, Vol. 160, No. 1, P. 106.

4. **Kreutz-Delgado K. et al.** Dictionary learning algorithms for sparse representation, *Neural computation*, 2003, Vol. 15, No. 2, Pp. 349–396.

5. **Pedregosa F. et al.** Scikit-learn: Machine learning in Python, *The Journal of Machine Learning Research*, 2011, Vol. 12, Pp. 2825–2830.

6. **Erhan D. et al.** Why does unsupervised pretraining help deep learning? *The Journal of Machine Learning Research*, 2010, Vol. 11, Pp. 625–660.

ХУРШУДОВ Артем Александрович – *аспирант кафедры информационных систем и программирования Института компьютерных систем и информационной безопасности Кубанского государственного технологического университета.*

350072, Россия, Краснодарский край, г. Краснодар, ул. Московская, д. 2.
E-mail: art1783@gmail.com

KHURSHUDOV, Artem A. *Kuban State Technological University.*
350072, Moskovskaya Str. 2, Krasnodar, Krasnodar krai, Russia.
E-mail: art1783@gmail.com

МАРКОВ Виталий Николаевич – *профессор кафедры информационных систем и программирования Института компьютерных систем и информационной безопасности Кубанского государственного технологического университета, доктор технических наук.*

350072, Россия, Краснодарский край, г. Краснодар, ул. Московская, д. 2.
E-mail: vinitar@yandex.ru

MARKOV, Vitaliy N. *Kuban State Technological University.*
350072, Moskovskaya Str. 2, Krasnodar, Krasnodar krai, Russia.
E-mail: vinitar@yandex.ru