

НЕЙРОСЕТЕВОЙ АЛГОРИТМ ОТСЛЕЖИВАНИЯ ДВИЖЕНИЯ ОБЪЕКТА НА МНОЖЕСТВЕ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ

Ярославцева Т.С. Email: Yaroslavtseva1133@scientifictext.ru

*Ярославцева Тамара Сергеевна – студент,
кафедра кибернетики и мехатроники,
Российский университет дружбы народов, г. Москва*

Аннотация: в данной статье описывается применение искусственных нейронных сетей для решения задачи отслеживания объекта. Исходные данные представлены последовательностью цифровых изображений. В качестве сети для обучения выбрана глубокая сверточная нейронная сеть с одним входным слоем, двумя сверточными и двумя субдискретизирующими. Выходом алгоритма являются вероятности соответствия той или иной части изображения искомому объекту. В результате выбирается кадр с максимальной вероятностью и принимается за положение объекта в данный момент. По итогам проведенного эксперимента точность отслеживания объекта составила 79%.

Ключевые слова: отслеживание объекта, сверточная нейронная сеть, глубокое обучение.

NEURAL NETWORK ALGORITHM TRACKING AN OBJECT ON A SET OF DIGITAL IMAGES

Yaroslavtseva T.S.

*Yaroslavtseva Tamara Sergeevna – student,
CYBERNETICS AND MECHATRONICS DEPARTMENT
PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA,
MOSCOW*

Abstract: this article describes the usage of artificial neural networks in the issue of tracking an object. Initial data is a set of digital images. The deep convolutional neural network with one input layer, two convolutional and two down subsampling layers is selected as a network for training. The results of the algorithm performed are probabilities of matching between parts of the original image and the required object. In consequence, the frame with the highest probability is selected; hence, it is taken as the position of the object at the moment. According to the conducted experiment, the tracking accuracy of the object has made up 79%.

Keywords: object tracking, convolutional neural network, deep learning.

УДК 331.225.3

Введение

Несмотря на большие достижения в изучении признаков изображения в различных визуальных компьютерных задачах, глубокие нейронные сети, как правило, рассматриваются как непригодные для визуального отслеживания, потому что они требуют очень много времени на обучение и большое количество обучающих примеров-образцов.

Признаки изображения играют решающую роль во многих сложных визуальных компьютерных задачах, таких как распознавание и обнаружение объекта. К сожалению, во многих алгоритмах, признаки определены и скомпонованы вручную [3]. Даже если эти методы доставляют удовлетворительные результаты на отдельных наборах данных, все же признаки, заданные вручную, будут ограничивать эффективность отслеживания. Например, нормированная взаимная корреляция - определяющий метод, при благоприятных световых условиях, может стать неэффективным, когда объект движется в тени. Это требует хороших механизмов обучения для визуального отслеживания, которые были бы способны захватывать внешние признаки значительно изменяющиеся во времени.

Недавно, глубокие нейронные сети получили значительное внимание благодаря своим достижениям в области изучения признаков визуальной информации. В отличие от традиционных признаков, заданных вручную [4], многослойная архитектура нейронных сетей может эффективно захватить сложные иерархии, описывающие исходные данные. В частности, сверточные нейронные сети (СНС) показали превосходную производительность на стандартных задачах распознавания объекта [7], что позволяет эффективно изучать сложные отображения, при этом используя минимальные области.

Тем не менее, процесс принятия СНС для отслеживания объекта не такая простая задача. Прежде всего, СНС требуется большое количество обучающих образцов. Кроме того, обучение СНС является напряженным вычислительным процессом.

В данной работе представлен алгоритм отслеживания объекта с помощью одной Сверточной Нейронной Сети (СНС).

Постановка задачи

Задача состоит в построении нейросетевого алгоритма, способного отслеживать объект на последовательности цифровых изображений $S = (S_1, S_2, \dots, S_n)$ и имеющего ошибку отслеживания не более чем на $p\%$ изображений. Формально этот процесс можно разбить на четыре этапа.

Первый этап состоит в отборе и подготовке множества данных. Подготовка заключается в применении некоторой функции преобразования $F: S_i \rightarrow S_i^*$ к каждому элементу упорядоченного множества S . Необходимость применения функции преобразования F вытекает из избыточности мощности множества признаков отслеживаемого объекта.

На втором этапе определяется архитектура и тип искусственной нейронной сети (ИНС). Поскольку для обработки изображений зачастую применяются сверточные нейронные сети (СНС), нам необходимо определить следующие параметры:

- Размерность входного слоя
- Размерность выходного слоя
- Количество слоев свертки
- Размерности слоев свертки
- Количество субдискретизирующих слоев
- Размерности субдискретизирующих слоев

Определение указанных параметров возможно либо с помощью опыта эксперта, либо при использовании алгоритмов метаоптимизации. Поскольку алгоритмы метаоптимизации чрезвычайно требовательны к вычислительным ресурсам, а СНС, в свою очередь, достаточно долго обучаются, то целесообразно использовать экспертный опыт.

Третий этап – обучение СНС. Обучение заключается в оптимизации множества матриц весовых коэффициентов $W = \{W_1, W_2, \dots, W_{m-1}\}$, где m – сумма количества слоев свертки и субдискретизирующих слоев. Для обучения используется один из градиентных алгоритмов. В процессе обучения используется подмножество $S' \subset S$. Из указанного подмножества формируются множества положительных и отрицательных примеров, причем к элементам S' предварительно применяется некоторая функция $g: S' \rightarrow s'$, такая что размерность $s' < S'$. Обозначим множества положительных и отрицательных примеров s'_+ и s'_- соответственно.

На финальном этапе построенная и обученная СНС тестируется на таком подмножестве S'' , что $S' \cap S'' = \emptyset$. Качество построенной СНС определяется в соответствии со следующим критерием:

$$J = \frac{1}{k} \sum_{i=1}^k \|x_i - x'_i\|_2,$$

где k – мощность множества S'' , $\|\cdot\|_2$ – евклидова норма, x_i – координаты левого верхнего угла объекта, x'_i – координаты левого верхнего угла найденного объекта.

Эксперимент и результаты

Программа разработана в программном комплексе Matlab. Для проведения эксперимента была установлена дополнительная библиотека DeepLearnToolbox, позволяющая работать со сверточными нейронными сетями [1].

Тестовый набор данных Tiger1 получен с сайта.[2] Он представляет собой набор из 354 полноцветных изображений формата JPEG размера 640×480.

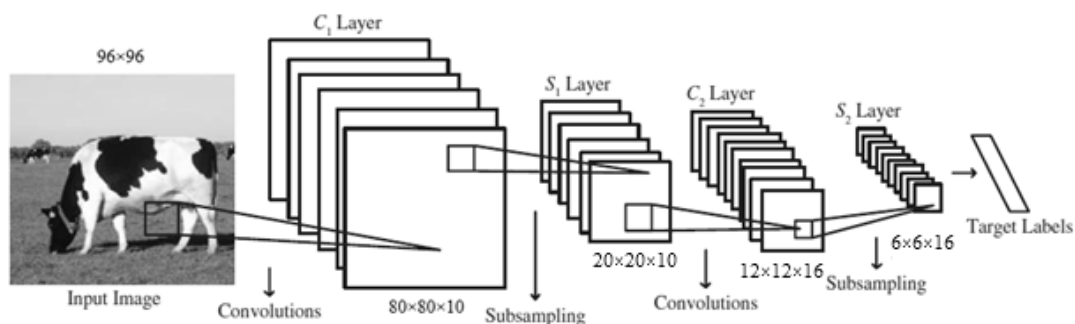


Рис. 1 Структура СНС

В качестве нейронной сети выбрана СНС с одним входным слоем, двумя сверточными и двумя субдискретизирующими. Размерность входного слоя 96×96 , первого сверточного слоя 80×80 , первого слоя подвыборки 20×20 . Данные слои состоят из 10-ти карт признаков. Размерности второго слоя свертки и субдискретизирующего слоя 12×12 и 6×6 соответственно. Они представлены 16-ю картами признаков. Структура сети представлена на

Рис. 1.

На начальном этапе работы программы цветные изображения конвертируются в изображения в градациях серого, а затем с помощью фильтров `imsharpen` и `imadjust`, увеличивающих резкость и усиливающих контуры, проводится предварительная обработка.

Следующим шагом является создание положительного и отрицательного набора обучающих и тестовых изображений размерности, соответствующей размерности входного слоя нейронной сети. Положительными считаются те изображения, которые содержат только изображение тигра, а все остальные – отрицательными.

Следующим этапом является обучение и тестирование СНС. Для этого экспериментальным путем определяются параметры сети: количество пакетов обучения **batchsize = 5** и константа обучения **$\alpha = 0.8$** .

Путем варьирования количества эпох и тестирования нашей сети на тестовой выборке определяем оптимальное число эпох. Минимальное количество ошибочных распознаваний соответствует числу эпох равному 25.

После обучения нейронной сети необходимо отследить объект на множестве исходных изображений. Для этого исходные изображения обходятся матрицей размерности 96×96 со сдвигом 48 пикселей и полученные кадры каждого изображения подаются на вход обученной нейронной сети.

На выходе получаются вероятности соответствия того или иного кадра изображению тигра на каждом из изображений. Выбирается кадр с максимальной вероятностью и принимается за положение тигра в данный момент.



Рис. 2. Результаты верного распознавания объекта



Рис. 3. Результаты ошибочного распознавания объекта

При работе программы количество ошибок составило 21%. То есть положение тигра не на каждом кадре определено верно.

Для вывода результатов на экран положение тигра выделяется на каждом кадре прямоугольником и полученная последовательность с выделенным объектом сохраняется в видеофайл. Результаты работы программы на некоторых кадрах представлены на

Рис. 2 и

Рис. 3.

Заключение

В представленной статье разработан нейросетевой алгоритм отслеживания движения объекта на множестве цифровых изображений. В качестве архитектуры ИНС выбрана сверточная нейронная сеть, которая обладает рядом преимуществ в распознавании изображений по сравнению с обычными нейросетями.

В результате работы программы выявлены следующие недостатки:

1. Центральный процессор неэффективен при работе с изображениями. Поскольку используемая библиотека работает с центральным процессором, а не с графическим обучением занимает большое количество времени.
2. Ручная подготовка обучающих примеров не позволяет создать достаточно большую базу для обучения
3. Существует большое количество искажений изображений и, в идеальном случае, СНС должна быть инвариантна ко всем. Однако при использовании только одного типа СНС это недостижимо.
4. Подбор параметров сети и алгоритма обучения выбирался экспертом, однако, метаоптимизационные алгоритмы возможно справились бы с задачей лучше.

В дальнейшем планируется провести данный эксперимент с привлечением больших мощностей, изменением структуры сети, а также оптимальным выбором ее параметров.

Список литературы / References

1. Библиотека изображений. [Электронный ресурс]. Режим доступа: <https://sites.google.com/site/trackerbenchmark/benchmarks/v10/> (дата обращения: 21.11.2016).
2. Библиотека СНС. [Электронный ресурс]. Режим доступа: <https://github.com/rasmusbergpalm/DeepLearnToolbox/> (дата обращения: 21.11.2016).
3. *Collins Robert T., Liu Yanxi, Leordeanu Marius.* Online selection of discriminative tracking features. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol. 27. № 10. Pp. 1631 – 1643, 2005.
4. *Dalal Navneet, Triggs Bill.* Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005.* Vol. 1. Pp. 886 – 893.
5. *Hanxi Li, Yi Li, Fatih Porikli, DeepTrack.* Learning Discriminative Feature Representations Online for Robust Visual Tracking. *CoRR abs. 1503.00072, 2015.*
6. *Jialue Fan, Wei Xu, Ying Wu and Yihong Gong.* Human Tracking Using Convolutional Neural Networks. *IEEE Transactions on Neural Networks.* Vol. 21. № 10. Pp. 1610 - 1623, 2010.
7. *Kavukcuoglu Koray, Gregor Karol, LeCun Yann.* Learning convolutional feature hierachies for visual recognition. *NIPS, 2010.*