

Источник: Інформатика та комп'ютерні технології / Матеріали VI міжнародної науково-технічної конференції студентів, аспірантів та молодих науковців – 23-25 листопада 2010 р., Донецьк, ДонНТУ. Т.2 – 2010, с. 120-123.
УДК 004.931

ПАРАМЕТРИЧЕСКАЯ МОДЕЛЬ АРХИТЕКТУРЫ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ

Умяров Н.Х., Федяев О. И.

Донецкий национальный технический университет

Задача распознавания образов относится к классу трудно формализуемых задач и в настоящее время является особенно актуальной в связи с необходимостью автоматизации образных процессов коммуникации (визуальных, речевых) в интеллектуальных системах. Поэтому до сих пор продолжается поиск и реализация эффективных принципов передачи распознавательной функции человека компьютеризированным системам. Для решения задач этого класса очень перспективны искусственные нейронные сети, как наиболее адекватные классу трудно формализуемых задач распознавания образов. Они индуцированы биологией, т. к. состоят из элементов, функциональные возможности которых аналогичны большинству элементарных функций биологического нейрона. Несмотря на поверхностное сходство, искусственные нейронные сети демонстрируют свойства, присущие живому мозгу. В частности, они обучаются на основе опыта, обобщают предыдущие прецеденты на новых примерах, извлекают существенные свойства из поступающей информации, содержащей лишние данные и т. д. [1].

В настоящее время предложено большое количество нейросетевых парадигм для решения задач распознавания образов. Значительные затруднения при распознавании вызывают образы, подверженные какому-либо искажению. В данной работе рассматривается новая модель нейросети из класса неокогнитронов - свёрточная нейронная сеть, которая успешно применяется для интеллектуальной обработки визуальных данных [3, 4]. Как известно, при решении задачи распознавания образов приходится сталкиваться с рядом проблем. Самая существенная из них – это сохранение и учёт топологии входного изображения. Пиксели можно подавать на вход в любом фиксированном порядке и это не повлияет на исход обучения. С другой стороны, образы имеют чёткую двухмерную структуру (соседние пиксели связаны между собой) и она несёт в себе ценную информацию о характере изображения. Локальная связь пикселей – основная причина необходимости использования новых механизмов извлечения локальных признаков на определенной подобласти изображения с последующим формированием некой системы таких признаков. Такая система является интуитивно понятной: различные конфигурации соседних пикселей формируют определенные категории (углы, стороны и т. д.). Очевидно, что эффективная система распознавания образов должна основываться на алгоритме, учитывающем такие особенности входных данных.

Другой проблемой является вариация представления одного и того же образа, а также его деформация (например, индивидуальный почерк, который существенно меняет написание символов). В принципе, эту проблему можно решить за счёт пополнения обучающей выборки примерами таких искажённых символов, однако это приведёт к низкой скорости обучения, а ещё хуже - к плохой обобщающей способности нейросети [2].

Решению этих проблем были посвящены работы американского учёного французского происхождения Яна Ле Куна, связанные с разработкой и исследованием свёрточных нейронных сетей [2, 5]. Идея свёрточных нейронных сетей заключается в чередовании свёрточных слоёв (C-layers), слоёв подвыборки (S-layers) и использовании

на выходе полносвязного слоя нейронов, которые в целом образуют ансамбль специализированных нейросетей. В основе свёрточных сетей лежат три механизма, используемых для достижения инвариантности к переносу, масштабированию, незначительным искажениям:

1) Локальное извлечение признаков. Каждый нейрон получает входной сигнал от локального рецептивного поля в предыдущем слое, извлекая, таким образом, его локальные признаки. Как только признак извлечён – его точное расположение уже не имеет значение, поскольку установлено его местонахождение относительно других признаков.

2) Формирование слоёв в виде набора карт признаков. Каждый вычислительный слой состоит из множества карт-признаков – плоскостей, на которых все нейроны должны использовать одно и то же множество синаптических весов. Такая форма усложняет структуру сети, однако имеет два важных преимущества: инвариантность к смещению, которое достигается с помощью свёртки с ядром небольшого размера, и сокращение числа свободных параметров, которое достигается за счёт совместного использования синаптических весов нейронами одной и той же карты.

3) Подвыборка. За каждым слоем свёртки следует вычислительный слой, осуществляющий локальное усреднение и подвыборку. За счёт этого достигается уменьшение разрешения для карт признаков. Такая операция приводит к понижению чувствительности выходного сигнала оператора отображения признаков к незначительному смещению и прочим видам деформации. В качестве такого оператора выступает одна из сигмоидальных функций, используемых при построении нейронных сетей (например, гиперболический тангенс) [2].

Следует заметить, что последовательное применение свёртки и подвыборки приводит к так называемому повышению уровня признаков: если первый слой извлекает локальные признаки из областей изображения, то последующие слои извлекают общие признаки, которые называются признаками высокого порядка. На рис. 1 показана сеть свёртки, реализующая распознавание входного изображений. Рассмотрим её архитектуру, особенности функционирования и параметрическое описание.

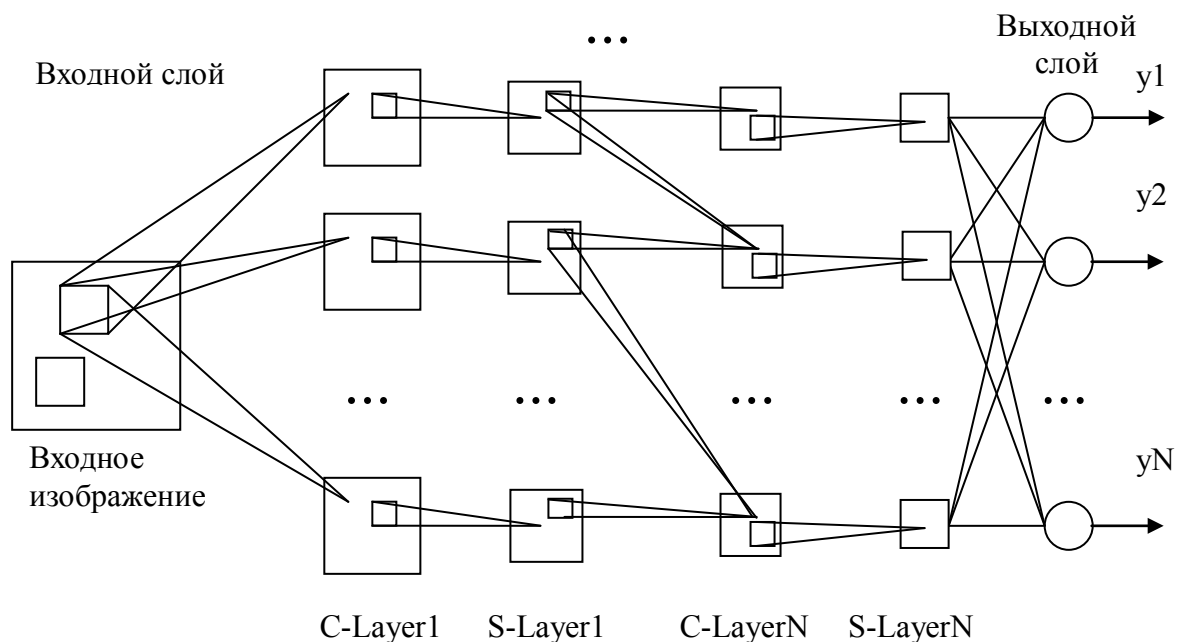


Рисунок 1 – Архитектура сверточной нейронной сети

Входное изображение лучше представлять двумерной матрицей пикселей. Это позволит сохранить топологию изображения при переходе по слоям работы сети и облегчить распознавание искажённое или повернутое изображение.

Основная особенность работы свёрточных сетей связана с использованием карт признаков. Изображение сканируется небольшим квадратным окном. Положения окна на изображении могут частично перекрываться. В квадрате окна каждому сканируемому пикселю сопоставлен свой вес и на выход окно выдаёт просуммированное взвешенное значение всех пикселей. Но поскольку положений окна много, то получается двумерная решётка таких значений. Получается, что это окно детектирует наличие какого-то сочетания пикселей (признака) на изображении. В итоге на выходной решётке получатся высокие значения там, где есть нужное сочетание пикселей, и низкие там, где содержание никак не похоже на то, что ищем. Эта решётка называется картой признаков. Причём находится не один признак, а много разных. Для каждого из них имеется своё окно со своим набором весов и в итоге получается несколько параллельных карт признаков на выходе. Каждая карта признаков (как изображение) подаётся на вход своего следующего слоя распознавателя, и тоже получает на выходе ещё несколько параллельных карт. В выходном слое нейросети все карты подаются на вход итогового распознавателя, в качестве которого можно использовать многослойный перцептрон [3, 4].

Для имитационного моделирования была составлена параметрическая модель архитектуры свёрточной нейронной сети. На теоретико-множественном уровне модель свёрточной нейронной сети CNN (Convolution Neural Network) представлена в виде кортежа

$$CNN = \langle M, \{N^\eta\}, \{ \langle \{C_{(kl),m}^{(\eta,p^\eta,(ij))}\}, f^\eta, \{S_{(qt),n}^{(\eta,p^\eta,(i',j'))}\}, \varphi^\eta \rangle \}, N, \{w_i^{j'}\}, \psi \rangle$$

где M – количество слоёв в сети, η – номер слоя $1 \leq \eta \leq M$; N^η – число карт признаков в η -м слое; $C_{(kl),m}^{(\eta,p^\eta,(ij))}$ – весовой коэффициент связи m -го входа (ij) нейрона p -й плоскости η -го слоя с (kl) -м выходом его рецептивной области (то есть вход нейрона – матрица рецептивных полей); $1 \leq i \leq I^{(\eta,c)}$, $1 \leq j \leq J^{(\eta,c)}$, $I^{(\eta,c)}$, $J^{(\eta,c)}$ – размеры карты признаков η -слоя C -го подслоя. $a_{ij}^{\eta,c} \leq k \leq b_{ij}^{\eta,c}$, $e_{ij}^{\eta,c} \leq l \leq d_{ij}^{\eta,c}$; $a_{ij}^{\eta,c}$, $b_{ij}^{\eta,c}$, $e_{ij}^{\eta,c}$, $d_{ij}^{\eta,c}$ – координаты границы области связи (ij)-го элемента нейрона η -слоя C -го подслоя. $1 \leq m \leq \theta$, $\theta = (b_{ij}^{\eta,c} - a_{ij}^{\eta,c}) \cdot (d_{ij}^{\eta,c} - e_{ij}^{\eta,c})$ – размер вектора входных сигналов (ij)-го нейрона η -слоя C -го подслоя.

Другие обозначения в кортеже означают: $S_{(qt),n}^{(\eta,p^\eta,(i',j'))}$ – весовой коэффициент связи n -го входа ($i'j'$) нейрона p -й плоскости η -го слоя с (qt) -м выходом его рецептивной области (то есть вход нейрона – матрица рецептивных полей); $1 \leq i' \leq I^{(\eta,s)}$, $1 \leq j' \leq J^{(\eta,s)}$, $I^{(\eta,s)}$, $J^{(\eta,s)}$ – размеры карты признаков η -слоя S -го подслоя; $a_{i'j'}^{\eta,s} \leq q \leq b_{i'j'}^{\eta,s}$, $e_{ij}^{\eta,s} \leq t \leq d_{ij}^{\eta,s}$; $a_{ij}^{\eta,s}$, $b_{ij}^{\eta,s}$, $e_{ij}^{\eta,s}$, $d_{ij}^{\eta,s}$ – координаты границы области связи ($i'j'$)-го элемента нейрона η -слоя S -го подслоя; $1 \leq n \leq \theta$, $\theta = (b_{ij}^{\eta,s} - a_{ij}^{\eta,s}) \cdot (d_{ij}^{\eta,s} - e_{ij}^{\eta,s})$ – размер вектора входных сигналов ($i'j'$)-го нейрона η -слоя S -го подслоя; f^η, φ^η – функции активации нейронов η -го слоя, C и S слоёв соответственно.

Опишем параметры выходного слоя CNN: N - число нейронов последнего слоя нейронной сети; $W_{i''}^{j''}$ - матрица весовых коэффициентов выходного слоя; ψ - функция активации нейронов выходного слоя.

В общем случае число слоёв может быть любое количество, однако Ян Ле Кун не рекомендует использовать слишком большое число слоёв, т. к. это не только не приводит к росту обобщающей способности сети (то есть способности распознавания большего числа примеров), а наоборот – к их ухудшению. Сравнительная характеристика однослойных, двухслойных и многослойных сетей приводится в [4].

Ещё одним преимуществом свёрточных сетей является их универсальность в задачах распознавания. Этот класс нейронных сетей можно использовать для распознавания лиц, проектирования систем компьютерного зрения. Применение свёрточных сетей вместе с классификатором Витерби позволяет достаточно качественно распознавать отдельные рукописные слова. Подвид свёрточных сетей, так называемые TDNN сети (Time Delay Neural Networks) применяются при распознавании речи. В этом случае операции свёртки и подвыборки выявляют определённые значения сигналов, а карты признаков привязаны к временной составляющей сигнала [6].

В заключении можно сказать, что применение свёрточных нейронных сетей является перспективным направлением в распознавании изображений, лиц и других образов, алгоритм распознавания которых трудно формализуем аналитически. Свёрточные сети лишены многих недостатков классического многослойного персептрона, обладают инвариантностью к позиционным изменениям изображения, шуму, сохраняют топологию изображения.

Литература

- [1] Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. – М.: Мир, 1992. – 260 с.
- [2] Le Cunn Y., Bottou L., Orr G. B. “Neural Networks: Tricks of the trade”, Springer. – 1998. – P. 1 – 5.
- [3] Федяев О.И., Куликов С.А. Многоуровневые нейронные модели типа неокогнитрон. Архитектура, обучение и распознавание // Труды 8-й Всероссийской конференции «Нейрокомпьютеры и их применение» НКП-2002. – М.: Век книги, 2002. С. 1046-1052.
- [4] Федяев О.И., Махно Ю.С. Система распознавания зашумлённых и искажённых графических образов на основе нейронной сети типа неокогнитрон // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008: Труды конференции. Т. 3. – М.: ЛЕНАНД, 2008. – 464 с.
- [5] Костецкая Г.Ю., Федяев О.И. Распознавание образов на основе свёрточных нейронных сетей // Материалы 5-ой всеукраинской научно-технической конференции студентов, аспирантов и молодых учёных КМИТ-2009. – Донецк, ДонНТУ, 2009. С. 274-276.
- [6] Le Cunn Y, Bengio Y., “Convolutional neural networks for Images, Speech and Time Series”, AT&T Laboratories, 1995, – P. 1 – 14.