

XV Международная научно-практическая конференция студентов аспирантов и молодых учёных
«Молодёжь и современные информационные технологии»

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОПРЕДЕЛЕНИЯ РАСПОЛОЖЕНИЯ КЛЮЧЕВЫХ ТОЧЕК ЛИЦА НА ИЗОБРАЖЕНИЯХ

О.И. Минаева

Научный руководитель: В.Г. Спицын, д.т.н., профессор кафедры ИСТ ИШИТР ТПУ
Томский политехнический университет
E-mail: olua94@mail.ru

Введение

Одной из самых сложных задач в информационных технологиях является обработка и распознавание объектов на изображениях.

Современные методы определения ключевых точек на изображениях лиц используются для решения широкого круга задач, а именно: распознавание лиц, отслеживание лиц на изображениях и видео, анализ эмоций человека, обнаружение дефектов лиц в медицинских диагнозах.

Определение ключевых лицевых точек является сложной проблемой, так как лицевые признаки сильно варьируются от человека к человеку, и даже для одного человека существует большое количество вариаций признаков в зависимости от позы, размера изображения, угла наклона головы и условий освещенности.

В данной статье рассматриваются нейронные сети, а также их подвид – сверточные нейронные сети, которые обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.

Нейронные сети

В настоящее время существует около десятка разновидностей нейронных сетей (НС). Одним из самых широко используемых вариантов является сеть, построенная на многослойном перцептроне, которая позволяет классифицировать поданное на вход изображение/сигнал в соответствии с предварительным обучением сети.

Обучаются нейронные сети на наборе обучающих примеров. Суть обучения сводится к настройке весов межнейронных связей в процессе решения оптимизационной задачи методом градиентного спуска. В процессе обучения НС происходит автоматическое извлечение ключевых признаков, определение их важности и построение взаимосвязей между ними.

Нейронная сеть принимает тренировочный набор «как есть» и учится производить

правдоподобное решение, не претендуя на абсолютную истину. То есть строится наилучшая нефизическая модель, которая не является максимально точным соответствием реального процесса, но даёт приемлемую его аппроксимацию.

Наилучшие результаты в области распознавания лиц показала сверточная нейронная сеть (СНС). Успех обусловлен возможностью учета двумерной топологии изображения, в отличие от многослойного перцептрона.

В основе сверточной нейронной сети (СНС) лежат идеи, предложенные Ле Куном и Бенджи [1]:

- локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов);
- общие веса (обеспечивают детектирование некоторых черт в любом месте изображения и уменьшают общее число весовых коэффициентов);
- иерархическая организация с пространственными подвыборками (subsampling).

На рисунке 1 можно увидеть схематичную архитектуру СНС.

Постановка задачи

В решаемой задаче входом модели является изображение, выходом – вычисленные определенным образом координаты ключевых точек лица. Такой набор координат можно представить в виде выходного вектора $(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$, где x_i – координата i -й точки по оси X , y_i – координата i -й точки по оси Y . Соответственно, решаемая задача формулируется как задача регрессии.

Для решения задачи необходимо подобрать архитектуру сети, с помощью которой можно получить наилучший результат. Результат будет оцениваться относительно рейтинга в соревновании Kaggle по соответствующей тематике [2].

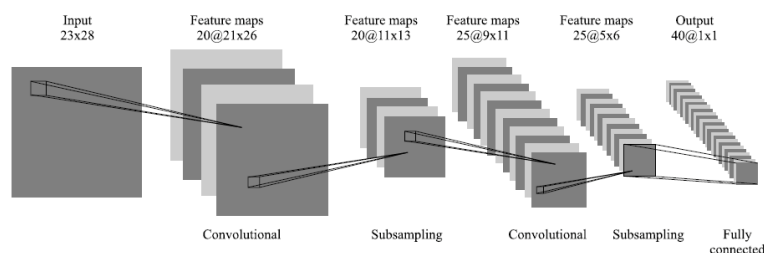


Рис. 1. Архитектура сверточной нейронной сети

Средства разработки

Для создания и настройки нейронных сетей в работе использовалась библиотека Theano с библиотекой-оберткой Lasagne [3, 4]. Выбранные библиотеки предоставляют набор готовых компонентов: слоёв, алгоритмов оптимизации, функций потерь, инициализаций параметров и т.д. Язык программирования – Python.

База изображений лиц Kaggle

Для обучения и тестирования сетей использовалась база Kaggle, содержащая изображения лиц размером 96x96 пикселей. Каждая предсказанная ключевая точка определяется парой значений (x, y) в пространстве индексов пикселей. Есть 15 ключевых точек, которые представляют некоторые элементы лица, например, центры левого и правого глаза, кончик носа и т.д. [2].

Реализация

В работе реализованы сети различных структур, как без использования сверточных слоев, так и с ними. Из всех тренировочных изображений базы Kaggle (7049 изображений) были отобраны только те, которые включают в себя полный набор ключевых точек (2140 изображений). Из них на тренировочную выборку идет 80%, на валидационную – 20%.

Валидационная выборка необходима для исключения переобучения сети. Явление переобучения подразумевает собой слишком точное запоминание сетью примеров тренировочного множества, при котором сеть теряет способность к обобщению [5].

Сеть без сверточных слоев

Сеть имеет следующую структуру.

Таблица 1. Структура сети

№	Название слоя	Количество нейронов в слое
1	Входной	9216
2	Полносвязный	100
3	Выходной	30

Как видно из таблицы 2, после 397-й эпохи средняя квадратичная ошибка на валидационной выборке начинает увеличиваться, поэтому для избежания переобучения сети, обучение прекращается.

Таблица 2. Ошибки сети без сверточных слоев

Эпоха	Среднеквадратичная ошибка	
	Тренировочная выборка	Валидационная выборка
397	0.00225	0.00333
398	0.00225	0.00345

Сеть со сверточными слоями

В работе были проведены эксперименты с несколькими архитектурами сверточных сетей. Далее приведена архитектура сети, показавшей наилучший результат.

Таблица 3. Структура сети

№	Название слоя	Количество нейронов в слое
1	Входной	1x96x96
2	Сверточный1	32x94x94
3	Подвыборки1	32x47x47
4	Сверточный2	64x46x46
5	Подвыборки2	64x23x23
6	Сверточный3	128x22x22
7	Подвыборки3	128x11x11
8	Полносвязный	500
9	Выходной	30

В данном случае после 400 эпох результат следующий (таблица 4).

Таблица 4. Ошибка сети со сверточными слоями

Эпоха	Среднеквадратичная ошибка	
	Тренировочная выборка	Валидационная выборка
399	0.00201	0.00223
400	0.00201	0.00223

Заключение

По результатам экспериментов было обнаружено, что применение слоев свертки уменьшает среднеквадратическую ошибку сети при одном и том же количестве эпох обучения.

Полученный результат занимает 44-е место из 175 в рейтинге соревнования Kaggle. Это является неплохим результатом, учитывая то, что в работе не были применены методы наращивания тренировочной выборки (data augmentation), метод dropout, метод предобучения сети и некоторые другие. Перечисленные способы уменьшения ошибки сети планируется применить в дальнейшей работе.

Список использованных источников

1. Le Cun Y., Bengio Y. Convolutional networks for images, speech and time series // The handbook of brain theory and neural networks. - 1998. - V. 7. - № 1. - P. 255-258.
2. Kaggle Facial Keypoints Detection – [Электронный ресурс]. Режим доступа - <https://www.kaggle.com/c/facial-keypoints-detection>. Дата обращения: 20.11.17.
3. Theano at a Glance – [Электронный ресурс]. Режим доступа - <http://deeplearning.net/software/theano/introduction.html#introduction>. Дата обращения: 20.11.17.
4. Lasagne. API Reference – [Электронный ресурс]. Режим доступа - <https://lasagne.readthedocs.io/en/latest/>. Дата обращения: 20.11.17.
5. Overfitting and Underfitting With Machine Learning Algorithms – [Электронный ресурс]. Режим доступа - <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/>. Дата обращения: 20.11.17.