Московский государственный технический университет им. Н.Э.

Баумана

Кафедра «Системы обработки информации и управления»

Домашнее Задание №1

по дисциплине

«Методы машинного обучения»

Выполнил:

студент группы ИУ5-24М

вань хао

Москва — 2022 г.

Задание

Домашнее задание по дисциплине направлено на анализ современных методов машинного обучения и их применение для решения практических задач. Домашнее задание включает три основных этапа:

- выбор задачи;
- теоретический этап;
- практический этап.

1. Выбор задачи

Моя тема-Распознавание человеческих эмоций с помощью конволюционных нейронных сетей. Классификация человеческих эмоций остается важной задачей для многих алгоритмов компьютерного зрения, особенно в эпоху гуманоидных роботов, которые сосуществуют с людьми в повседневной жизни. Предлагаемые в настоящее время методы распознавания эмоций используют многослойные конволюционные сети для решения этой задачи, без явного вывода каких-либо признаков лица на этапе классификации.

На рисунке 1 показаны шесть основных человеческих выражений: счастье, грусть, страх, гнев, отвращение и удивление. Наша цель - определить и классифицировать эти шесть выражений на картинке.

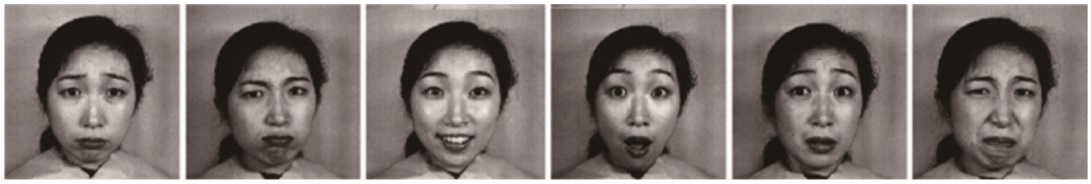

рис1-Шесть основных выражений

2. Теоретический этап

Тема найденной статьи: « Classifying and Visualizing Emotions with Emotional DAN» - «Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution».

На примере распознавания выражения лица в статье показано, как глубокие сверточные нейронные сети (DCNN) могут обучаться на основе зашумленных меток. Более конкретно, для каждого входного изображения имеется 10 маркировщиков, и сравниваются четыре различных подхода к использованию нескольких меток: мажоритарное голосование, обучение по нескольким меткам, вероятностное рисование меток и потеря перекрестной энтропии. Результаты показывают, что традиционная схема голосования по большинству голосов работает не так хорошо, как два последних

метода, которые полностью используют распределение меток. Кроме того, расширенный набор данных FER+, в котором каждое изображение лица имеет несколько меток, также будет предоставлен исследовательскому сообществу.

В первой статье были представлены соответствующие понятия, касающиеся модели DAN.DAN основана на архитектуре сети глубокого выравнивания, которая первоначально была предложена для надежного выравнивания лиц. Основное преимущество DAN перед другими конкурирующими методами выравнивания лица заключается в итерационном процессе корректировки положения лицевых ориентиров. Итерационный процесс встроен в архитектуру нейронной сети, поскольку с помощью тепловой карты ориентиров лица информация о местоположении ориентиров, обнаруженных на предыдущем этапе (слое), передается на следующий этап. Таким образом, в отличие от других конкурирующих методов, DAN может обрабатывать все изображение лица, а не его фрагменты, что приводит к значительному уменьшению несоответствий положения головы и улучшает его производительность в задачах распознавания ориентиров.

Модель сетевой архитектуры DAN показана на рисунке 2.

Name	Input shape	Output shape	Kernel
conv1a	224×224×1	224×224×64	3×3,1,1
conv1b	224×224×64	224×224×64	3×3,64,1
pool1	224×224×64	112×112×64	2×2,1,2
conv2a	112×112×64	112×112×128	3×3,64,1
conv2b	112×112×128	112×112×128	3×3,128,1
pool2	112×112×128	56×56×128	2×2,1,2
conv3a	56×56×128	56×56×256	3×3,128,1
conv3b	56×56×256	56×56×256	3×3,256,1
pool3	56×56×256	28×28×256	2×2,1,2
conv4a	28×28×256	28×28×512	3×3,256,1
conv4b	28×28×512	28×28×512	3×3,512,1
pool4	28×28×512	14×14×512	2×2,1,2
fc1	14×14×512	1×1×256	-
fc2_landmark	1×1×256	1×1×136	-
fc2_emotion	1×1×256	1×1×{3,7}	-

рисунке -2 Модель сетевой архитектуры

Первоначально DAN была вдохновлена системой каскадной регрессии формы, и аналогично она начинается с начальной оценки формы лица, которая уточняется после следующих итераций. В DAN каждая итерация представлена одним этапом глубокой нейронной сети. На каждом этапе (итерации) признаки извлекаются из всего изображения, а не из локальных участков изображения (в отличие от CSR).

Обучение состоит из последовательных этапов, где один этап состоит из нейронной сети с прямой передачей и соединительных слоев, генерирующих входные данные для следующего этапа. Каждый этап принимает три типа входных данных: входное изображение, выровненное по канонической форме, изображение особенностей, сгенерированное из плотного слоя предыдущего этапа, и тепловая карта ориентиров. Поэтому выход на каждом этапе DAN определяется как:

$$S_t = T_t^{-1}(S_{t-1}) + \Delta S_t,$$

где ΔSt - выход ориентиров на этапе t, а Tt - преобразование, используемое для приведения входного изображения к канонической позе.

В данной работе мы предполагаем, что способность DAN обрабатывать изображения с большой вариативностью и предоставлять надежную информацию об ориентирах на лице хорошо подходит для задачи распознавания эмоций. На Для этого мы расширяем задачу обучения сети с дополнительной целью оценки выраженных лицевых эмоций. Мы воплощаем эту идею, модифицируя функцию потерь с помощью суррогатного члена, который предназначен именно для задачи распознавания эмоций, и

минимизируем оба члена - по расположению ориентиров и по распознаванию эмоций - совместно. Таким образом, результирующая функция потерь L может быть выражена как:

$$\mathcal{L} = \alpha \cdot \frac{\| S_t - S^* \|}{d} - \beta \cdot E^* \cdot log(E_t),$$

где St - преобразованный результат прогнозирования лицевых ориентиров для этапа t, E - softmax результат прогнозирования настроения. S∗ - вектор местоположений наземных ориентиров, d - расстояние между наземными ориентирами как нормализованный скаляр, а E∗ - наземная истина метки настроения. Влияние каждого термина взвешивается с помощью коэффициентов α и β.

3. Практическая часть

Практическая часть выложена в Gitlab.

4 Заключение

Изучив методы распознавания эмоций и исследовав модель DAN, в данной работе мы описываем расширение нашего предыдущего подхода к распознаванию эмоций, которое позволяет использовать маркеры лица. Хотя результаты, полученные на наборе данных JAFFE, показывают, что еще есть возможности для улучшения, мы считаем,

что этот метод имеет большой потенциал, чтобы выйти за рамки предложенного в настоящее время подхода. Поэтому в дальнейшей работе мы сосредоточимся на улучшении нашего метода, используя механизм внимания на лицевых маркерах и экспериментируя с дополнительными условиями функции потерь. Мы также планируем исследовать другие применения нашего метода, например, в контексте детей с аутизмом, связанные с распознаванием эмоций.

5 Список использованных источников

1. Martin Abadi, Paul Barham, Jianmin Chen, ZhifengChen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. Tensorflow: A system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pages 265–283.

2.Kyunghyun Cho, Bart van Merri ̈enboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on

Empirical Methods in Natural Language Processing (EMNLP), pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

3. Kasra Hosseini, Federico Nanni Mariona Coll Ardanuy. 2020. DeezyMatch: A Flexible Deep Learning Approach to Fuzzy String Matching. EMNLP. Pages 62-69.