DOI: 10.34220/MAMSP_89-94
УДК 004.75

ТЕХНОЛОГИЯ ГЛУБОКОГО ОБУЧЕНИЯ В РАСПОЗНАВАНИИ ОБРАЗОВ

О.В. Оксюта[1], Сюй Ле[1], Р.С. Лопатин[1]

[1]ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова»

В статье рассматриваются методы распознавания лиц на основе сверточных нейронных сетей, проблемы распознавания лиц при наличии помех или замаскированности лица, рассмотрены основные этапы обучения нейронных сетей и процесса фактического распознавания.

Ключевые слова: распознавание лиц, проблемы обнаружения лиц, проблемы замаскированности, алгоритмы распознавания лиц, сверточные нейронные сети.

DEEP LEARNING TECHNOLOGY IN PATTERN RECOGNITION

O.V. Oksyuta, Xu Le, R.S. Lopatin

[1]Voronezh State University of Forestry and Technologies named after G.F. Morozov

The article discusses the methods of face recognition based on convolutional neural networks, the problems of face recognition in the presence of interference or face masking, the main stages of training neural networks and the process of actual recognition.

Keywords: face recognition, face detection problems, masking problems, face recognition algorithms, convolutional neural networks.

В последние годы с учетом потребностей общества в безопасности, защите жизни и защите информации, в разной степени были изучены различные технологии распознавания лиц, применяемые в различных сценариях. По сравнению с более ранними способами ручного выбора функций, на данном этапе стало традиционным обучать сети извлечению признаков, используя преимущества снижения арифметической мощности. Большая часть методов совре-

менного глубокого обучения, основанного на алгоритмах распознавания лиц, применяются к сценам без помех, когда лицо находится в более идеальной среде, изображение лица относительно четкое, без или с небольшим количеством декораций. Эти методы позволяют достичь высокой степени распознавания лиц при таких условиях [2].

Основными трудностями замаскированного распознавания лиц являются проблемы, связанные с маскировкой и межклассовой вариативностью. Алгоритмы замаскированного распознавания лиц должны иметь дело со сложными сценариями, которые рассматривают две вариации: одна – это человек, пытающийся запутать свою личность так, чтобы система не могла ее распознать; другая – это человек, имитирующий чужую личность.

На рисунке 1 показаны две группы изображений, a и b. В группе a два человека разные, но человек с левой стороны одет так, чтобы имитировать правое изображение, что делает их очень похожими. В группе b один и тот же человек, но использование аксессуаров, таких как солнечные очки и ковбойская шляпа справа, затрудняет их различение.

Сочетание этих двух сценариев, оба из которых могут привести к неправильной идентификации с помощью системы распознавания, само по себе делает замаскированное распознавание лиц трудной задачей. [4]

Второй проблемой обнаружения лиц является вероятность сбоя при обнаружении замаскированных изображений лиц, которая значительно выше, чем при обнаружении незамаскированных изображений лиц.

У этой проблемы есть две основные причины этого: во-первых, макияж или одежда человека делает цвет и текстуры похожи на окружающую среду, а вторая заключается в том, что человек носит аксессуары, которые могут затемнить большие детали лица. Такие ситуации делают алгоритмы обнаружения положение лица менее точными, или даже если положение лица определено, они не могут получить точки для выравнивания лица.

а) б)

Рисунок 1 – Сценарии замаскированного распознавания лиц

Эти проблемы решаются алгоритмами распознавания лиц на основе многозадачных конволюционных (сверточных) нейронных сетей – MNCNN (Multi-task Convolutional Neural Networks).[5] MTCNN – это трехступенчатый алгоритм обнаружения, который позволяет не только получить области лица, но и координаты пяти ключевых точек лица (глаза, кончик носа и углы губ). Структура MTCNN определяет три сети со схожими структурами P-Net, R-Net и O-Net [3].

Обучение. Так как методы обучения в трех сетях практически одинаковы, разница заключается только в различных размерах входных изображений. В качестве примера возьмем P-Net. Для изображения размером 12×12×3. Выход P-Net является признаком 1×1×16, который можно рассматривать как признак вектора u размерности 16. Первые два бита (u1, u2) используются для вычисления вероятности присутствия и отсутствия лица $p1$ и $p2$. [1]

Следующая функция потери кросс-энтропии используется непосредственно в обучении:

$$L_{\text{det}} = -y \cdot \log(p_1) - (1 - y) \cdot \log(p_2) \quad (1)$$

где y – характеристика изображения.

Четырехмерный вектор с третьего по шестой бит в изображении ($u3$, $u4$, $u5$, $u6$) являются координатами граничного поля. Координаты левой верхней и правой нижней точек не используются для обозначения ограничивающего поля, для обозначения ограничивающего поля используются координаты центральной точки, а также ширина и высота ограничивающего поля. MTCNN использует координаты левой верхней точки выходного ограничивающего окошка, а также высоту и ширину (l, t, h, w) окошка для отображения положения ограничивающего окошка. Для приближения полученного ограничивающего окошка к реальному помеченному ограничивающему окошку используется регрессионная функция. Функция потери вычисляется по формуле (2):

$$L_{box} = \| \boldsymbol{u}^{box} - \hat{\boldsymbol{u}}^{box} \|_2^2 \quad (2)$$

где $\boldsymbol{u}^{box}$ - четырехмерный вектор вывода сети, а $\hat{\boldsymbol{u}}^{box}$ – это вектор маркировки. Последние десять бит вектора $\boldsymbol{u}$ – это координаты пяти ключевых точек на лице человека. Получение координат ключевых точек считается регрессионной задачей. Проблема регрессии связана с использованием функции потери, основанной на евклидовом расстоянии:

$$L_{\text{landmark}} = \| \boldsymbol{u}^{\text{landmark}} - \hat{\boldsymbol{u}}^{\text{land mark}} \|_2^2 \quad (3)$$

где $\boldsymbol{u}^{\text{landmark}}$ - вектор координат ключевых точек сетевого вывода, $\hat{\boldsymbol{u}}^{\text{landmark}}$ – это вектор координат.

Данные по обучению обычно делятся на четыре категории: положительные пробы, отрицательные пробы, частичные образцы и образцы ключевых точек. Обучающие данные обычно получают путем перехвата небольших кусочков вокруг части изображения, содержащей лицо, с помощью скользящего окна. Вычисляется соотношение площадей перехваченной части и ограничивающего окошка. Если значение отношения больше 0,65 - положительная выборка; если меньше 0,3 - отрицательная выборка; если больше 0,4, но меньше 0,65 - частичная выборка. Выборка ключевых точек содержит только координаты ключевых точек [6].

Для четырех типов учебных проб используются положительные и отрицательные пробы, чтобы обучить сеть способности различать лица; используются положительные пробы и частичные образцы для обучения регрессии ограничивающих рамок; и образцы ключевых точек для обучения обнаружению ключевых точек. В связи с необходимостью одновременного обучения нескольким задачам, входные данные состоят из четырех типов обучающих выборок в соотношении 3:1:1:2 (отрицательные выборки: положительные выборки: частичные выборки: выборки ключевых точек). Общая потеря рассчитывается путем наложения трех описанных выше функций потери. MTCNN использует следующую функцию полной потери.

$$L = \sum_{i=1}^{n} \sum_{j \in \{\text{det,box,landmark}\}} \alpha_j \beta_j^i L_j^i \qquad (4)$$

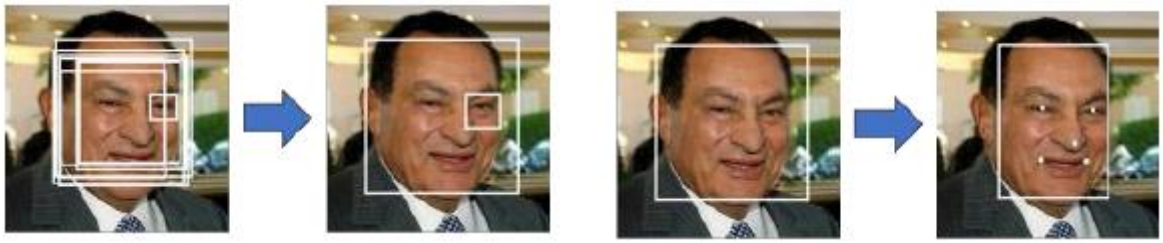
где α_jобозначает вес каждой функции потери при суммировании, β_j^iобозначает i-е изображение в каждой партии входного изображения, установленном как для обучения; для каждого изображения в наборе данных есть три вида задач β_{det},β_{box} и $\beta_{landmark}$, которыеобозначают классификационную задачу, обнаружение ограничивающих рамок и определение ключевых точек соответственно, а значение 1 указывает на то, что задача является необходимой, а 0 - нет. Для положительных, отрицательных, частичных выборок и выборок ключевых точек (β_{det}, β_{box}, $\beta_{landmark}$) являются (1,1,0), (1,0,0), (0,1,0) и (0,0,1).

Шаги, выполняемые во время обнаружения, приведены на рисунке 2. Так как P-Net является полностью конволюционной сетью, она может вводить данные любого размера изображения. Ограничительные рамки и ключевые точки в каждом блоке могут быть сведены к координатам исходного растра таким обра-

зом, чтобы можно было получить ограничивающие рамки и ключевые точки исходного растра. Проблема заключается в том, что лица в наборе данных, используемом для обучения P-Net, как правило, ограничены 12 × 12 изображениями, которые обычно занимают всего несколько десятков изображений. Но диапазон лиц на исходном изображении может быть значительно больше этого числа. Небольшой блок может соответствовать лишь небольшой части лица, что приводит к ошибке детектирования. Для решения этой проблемы MTCNN использует метод пирамиды изображений для многомасштабированного уменьшения исходного входного изображения. Затем изображения вводятся в P-Net, и каждый масштаб входного изображения определяет только лица, соответствующие этому масштабу. После результирующие граничные поля масштабируются и отражаются обратно к исходному изображению, так что на выходе P-Net получаются все граничные поля. Затем остальные ограничивающие поля подвергаются не максимальному подавлению (NMS).

Предварительная обработка: Изобразительная пирамида

1. NMS обработка вывода P-Net



2.Обработка NMS вывода R-Net

3.NMS обработка вывода O-Net

Рисунок 2 – Этапы процесса распознавания лиц

Эффект NMS заключается в удалении обнаруженных кадров с высоким перекрытием, и процесс происходит следующим образом:

1) каждому ограничивающему полю присваивается оценка, а в MTCNN оценка получается, как вероятность присутствия или отсутствия лица;

2) сортировка этих ограничивающих полей по баллам от наибольших до наименьших;

3) выбирается первое ограничивающее поле в последовательности (с наибольшей оценкой) и помещается в очередь на удержание, затем вычисляется IOU с последующими ограничивающими по очереди полями и определяется, превышает ли она пороговое значение, где пороговое значение может быть принята за 0,7;

4) ограничивающие поля, превышающие пороговое значение, отбрасываются, тогда оставшиеся ограничивающие поля сортируются по баллам в порядке убывания (оставшиеся не включают те, которые находятся в очереди на удержание) и происходит возврат к шагу 3.

Текущий алгоритм распознавания лиц должен опираться на большое количество данных для обучения, чтобы получить определенные результаты, но это значительно уступает использованию специализированного маскировочного массива данных о лицах большой емкости, так что данные могут собираться в сети, чтобы получить лучшие обучающие ресурсы.

Список литературы

1. Дуда, Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт. – Санкт-Петербург : Лань, 2016. – 164 с.

2. Змеев, А.А. Сравнительный анализ архитектур нейронных сетей для использования их на практике / А.А. Змеев, В.В. Лавлинский, С.Н. Яньшин // Моделирование систем и процессов. – 2017. – Т. 10, № 4. – С. 18-26.

3. Лавлинский, В.В. Применение математического описания действий для целенаправленных систем на основе методов нейронных сетей / В.В. Лавлинский, С.Н. Яньшин // Моделирование систем и процессов. – 2017. – Т. 10, № 2. – С. 17-23.

4. Модификация метода поиска информации в сети интернет на основе использования методов индуктивного рассуждения / В.В. Лавлинский, А.Л. Савченко, И.А. Земцов, О.Г. Иванова // Моделирование систем и процессов. – 2019. – Т. 12, № 1. – С. 61-67.

5. Оксюта, О.В. Система распознавания дорожных знаков с использованием искусственных нейронных сетей / О.В. Оксюта, А.М. Милютин // Моделирование систем и процессов. – 2017. – Т. 10, № 1. – С. 64-67.

6. Taigman, Y. Deepface: Closing the gap to human-level performance in face verification[C] / Y. Taigman, M. Yang, M.A. Ranzato // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2014. – Pp/ 1701-1708. – DOI: 10.1109/CVPR.2014.220