

РАСПОЗНАВАНИЕ ОБРАЗОВ С ИСПОЛЬЗОВАНИЕМ СИАМСКОЙ НЕЙРОННОЙ СЕТИ

В работе описываются результаты исследования сиамской нейронной сети. Предлагаемый подход предлагается как альтернатива сверточной нейронной сети, в условиях ограниченного объёма обучаемых данных. Также открыт вопрос использования данного подхода в системах реального времени, в системах распознавания лиц и аутентификации. Описанию исследований этого подхода и посвящён данный тезис.

ВВЕДЕНИЕ

В системах распознавания лиц мы хотим, чтобы мы могли распознавать человека посредством одной фотографии. В случае, если система не может распознать изображение, это означает, что изображение человека не сохраняется в базе данных. Для решения данной проблемы использование только сверточной нейронной сети не возможно, поскольку сверточные нейронные сети не работают на малом тренировочном наборе, а также необходимо обучать модель снова, когда в системе появляется новый человек.

I. СИАМСКИЕ НЕЙРОННЫЕ СЕТИ

Для решения данной проблемы можем использовать сиамскую нейронную сеть. Главной целью сиамской нейронной сети является определение, насколько похожи два изображения.

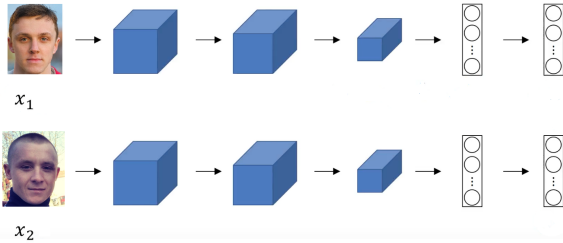


Рис. 1 – Преобразование входных изображений в векторы x_1 и x_2

Чтобы сравнить два изображения, необходимо вычислить расстояние d между векторами этих изображений.

$$d(x_1, x_2) = \|f(x_1) - f(x_2)\|^2 \quad (1)$$

Если это расстояние меньше порогового значения, это означает, что на изображениях одно и то же лицо.

II. ВЫВОД МАТЕМАТИЧЕСКОЙ МОДЕЛИ РАСПОЗНАВАНИЯ ОБРАЗОВ

Для того, чтобы получить конечный вектор из входного изображения, необходимо узнать входные параметры для этого изображения. Для

этого применим градиентный спуск к функции потери триплета. Используя три изображения: корневое (A), позитивное (P , тот же человек, что и корень), а также негативное (N), необходимо найти расстояния $d(A, P)$, и чтобы оно было меньше или равно расстоянию $d(A, N)$. Функция потери триплета минимизирует расстояние между изображением A и P , и максимизирует расстояние между изображением A и N . В этом случае модель может научиться создавать одинаковое кодирование для разных изображений, что означает, что расстояние между разными изображениями будут равны нулю. Для решения данной проблемы добавим α (вес), и чтобы всегда существовал разрыв между позитивным и негативным изображением.

$$d(A, P) + \alpha \leq d(A, N) \quad (2)$$

Вид функция потери триплета:

$$L(A, P, N) = \max(d(A, P) - d(A, N) + \alpha, 0) \quad (3)$$

Тогда функция стоимости – это сумма всех индивидуальных потерь в разных триплетах из всего набора тренировок.

$$COST = \sum_{i=1}^n L(A_i, P_i, N_i) \quad (4)$$

Тренировочный набор должен включать несколько изображений одного и того же человека, чтобы иметь пары A и P , а затем, после обучения, появится возможность распознавать людей посредством одной фотографии.

III. ВЫВОДЫ

Предлагаемый подход по распознаванию образов позволяет точно распознавать лица на фотографии. Данный подход может использоваться в системах реального времени, распознавания людей на фотографиях, аутентификации.

1. Advance AI: Face recognition using Siamese networks [Электронный ресурс]. – <https://mc.ai/advance-ai-face-recognition-using-siamese-networks>

Буяльский Олег Андреевич, магистрант кафедры систем управления БГУИР, oleg.buyalski@gmail.com.