

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Национальный исследовательский университет «МЭИ»

Институт: ИВТИ Кафедра: ВМСС
Направление подготовки: 09.03.01 Информатика и вычислительная техника

ОТЧЕТ по практике

**Наименование
практики:**

Производственная практика: технологическая (проектно-технологическая) практика

СТУДЕНТ

_____/ Ветошкин Л.Н. /
(подпись) (Фамилия и инициалы)
Группа А-08-19
(номер учебной группы)

**ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ
ПО ПРАКТИКЕ**

(отлично, хорошо, удовлетворительно, неудовлетворительно)
_____/ Кружилов И.С. /
(подпись) (Фамилия и инициалы члена комиссии)
_____/ _____ /
(подпись) (Фамилия и инициалы члена комиссии)

**Москва
2022**

Содержание отчета

ЗАДАНИЕ НА ПРАКТИКУ	3
ГРАФИК ПРОХОЖДЕНИЯ ПРАКТИКИ	4
Введение	6
Основная часть	7
1.1. Задача оценки двумерной позы человека	7
1.1.1 Stacked Hourglass Networks for Human Pose Estimation	7
1.1.2 PyraNet. Learning Feature Pyramids for Human Pose Estimation	9
1.3 Обработка тепловых карт	12
1.3.1 Применение argmax к тепловой карте	12
1.3.2 Нахождение «центра тяжести» тепловой карты	12
1.3.3. Сдвиг ко второму максимальному значению	12
1.3.4. Оценка с помощью функции максимального правдоподобия	13
2.1 Используемые данные	14
2.1.1 Датасет МРП Human Pose	14
2.1.2 Аугментация данных	14
3.1 Метрика	16
4.1 Результаты	16
Заключение	18
Список литературы	19

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Национальный исследовательский университет «МЭИ»

Институт: ИВТИ Кафедра: ВМСС
Направление 09.03.01 Информатика и вычислительная
подготовки/специальность: техника

ЗАДАНИЕ НА ПРАКТИКУ

Наименование практики: Производственная практика: технологическая (проектно-технологическая) практика
Студент: Ветошкин Лука Николаевич
(Фамилия, имя, отчество (при наличии) полностью)
Группа: А-08-19
(номер учебной группы)
Место прохождения практики: Институт информационных и вычислительных технологий, кафедра ВМСС
(наименование предприятия, организации, учреждения, подразделения МЭИ в соответствии с приказом о направлении на практику)
Сроки практики: 14.06.22 по 11.07.22 г.
(в соответствии с приказом о направлении на практику)

Содержание задания:

1. Оценка позы человека по фотоснимку с помощью нейросети
2. Обработка тепловых карт
3. Оценка полученных результатов
- 4.
- 5.

(вопросы, подлежащие изучению в соответствии с планируемыми результатами обучения, заполняются руководителем практики от МЭИ)

По результатам прохождения практики студент оформляет отчет по установленной форме.

Руководитель практики

/ Кружилов И.С. /
(подпись) (Фамилия и инициалы)

Студент

/ Ветошкин Л.Н. /
(подпись) (Фамилия и инициалы)

ГРАФИК ПРОХОЖДЕНИЯ ПРАКТИКИ

Номер п/п	Перечень работ в соответствии с заданием	Отметка о выполнении работы (выполнено /не выполнено)
1	<i>Обзор релевантной литературы</i>	выполнено
2	<i>Построение нейросетевой архитектуры</i>	выполнено
3	<i>Обработка тепловых карт</i>	выполнено
4	<i>Оценка полученных результатов</i>	выполнено
5		

Руководитель практики

/ **Кружилов И.С.** /

(подпись) (Фамилия и инициалы)



ХАРАКТЕРИСТИКА – ОТЗЫВ ПО ПРАКТИКЕ

*Производственная практика:
технологическая (проектно-
технологическая) практика*

(Наименование практики)

Студент: Ветошкин Л.Н.
(Фамилия, И.О.)

Группа: А-08-19
(номер учебной группы)

Место
прохождения
я практики: *Институт информационных и
вычислительных технологий*

(наименование предприятия, организации, учреждения,
подразделения МЭИ в соответствии с приказом о
направлении на практику)

Сроки
практики: 14.06.22 - 11.07.22 г.

(в соответствии с приказом о направлении на
практику)

Характеристика работы студента:

(В правом столбце проставьте **V** (галочку) в соответствующей строке)

№	Наименование критерия оценивания	Качественная оценка работы студента на практике	Отметка
1	Студент получил задание и приступил к его выполнению	своевременно	V
		с запозданием	
		с большим запозданием	
2	Полнота и целостность выполнения задания	работа выполнена полностью в соответствии с заданием	V
		отдельные разделы не согласованы между собой	
		задание выполнено фрагментарно	
3	Отношение студента к выполнению задания	отличное	V
		хорошее	
		удовлетворительное	
4	Готовность к решению поставленных задач по данной тематике	высокая	V
		средняя	
		низкая	
5	Соответствие отчетности требованиям	соответствует	V
		не соответствует	
6	За выполненную работу студент заслуживает оценку	отлично	V
		хорошо	
		удовлетворительно	

Работа студента в период прохождения практики выполнена в соответствии с заданием.

Студент Ветошкин Л.Н. допускается к промежуточной аттестации.
(Фамилия и инициалы)

Руководитель практики (от МЭИ)

/ Кружилов И.С. /
(подпись) (Фамилия и инициалы)

Введение

Задача оценки позы человека по фотоснимкам имеет все большее внимание со стороны научного сообщества. Так как выявляя ключевые точки на теле человека, можно научить компьютеры, роботов взаимодействовать с человеком; применять в видеоиграх; анализировать движения спортсменов; контроль безопасности и т.д. Из-за успешного применения глубокого обучения в компьютерном зрении, также можно применять его в задаче оценки позы.

В данной работе производится обзор релевантной литературы, была построена нейросетевая архитектура, для локализации точек скелета, были разобраны разные способы обработки тепловых карт.

Для обучения нейросети использовался язык программирования Python с фреймворком Pytorch. Исходный код проекта опубликован в репозитории на github (<https://github.com/richlukich/Keypoints>)

Основная часть

1.1. Задача оценки двумерной позы человека

Оценка позы человека является задачей локализацией человеческих суставов (локти, запястья, колени и т.д.) на изображениях или видео. Также эту задачу относят к поиску специфичных поз в пространстве всех поз. В этой работе приведено решение задачи оценки позы в 2D, то есть будут оцениваться координаты (x,y) суставов по фотографии для одного человека. Также существует задачи оценивания позы в 3D, когда оцениваются 3 координаты (x,y,z) , и для нескольких людей на одной фотографии.

1.1.1 Stacked Hourglass Networks for Human Pose Estimation

Данная архитектура появилась в 2016 году и показала очень хорошее качество по сравнению с другими архитектурами. К тому же она является интуитивно понятной, что является также большим плюсом.

Авторы статьи Stacked Hourglass Networks for Human Pose Estimation [1] предлагают использовать идею каскадной архитектуры с несколькими стадиями улучшений предсказаний. Общая схема сети изображена на рисунке 1.1.

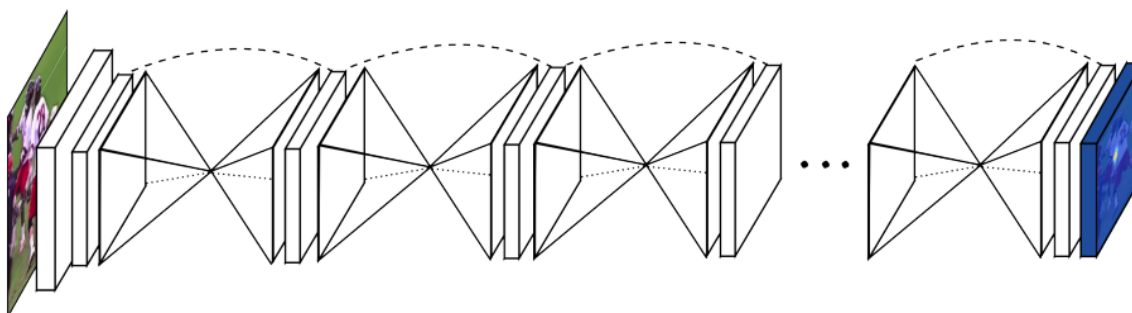


Рис 1.1 Stacked Hourglass Network

Данная сеть основана на каскадировании «hourglass modules» (рисунок 1.2), которые представляют собой полносверточную сеть и сделаны для того, чтобы обрабатывать информацию в разных масштабах. Она использует технологию блоков «residual block» из ResNet [2] и технологию «кодировщик-декодировщик» из архитектуры U-Net [3] предназначенной для решения задач сегментации. На каждом этапе трансформации создается отдельная ветвь, в которой сохраняется информация о текущем разрешении. Трансформация состоит из применения «residual module» и слоя субдискретизации, когда будет достигнуто разрешение 4x4. Трансформация будет аналогичная, только вместо слоя субдискретизации будет слой повышения дискретизации.

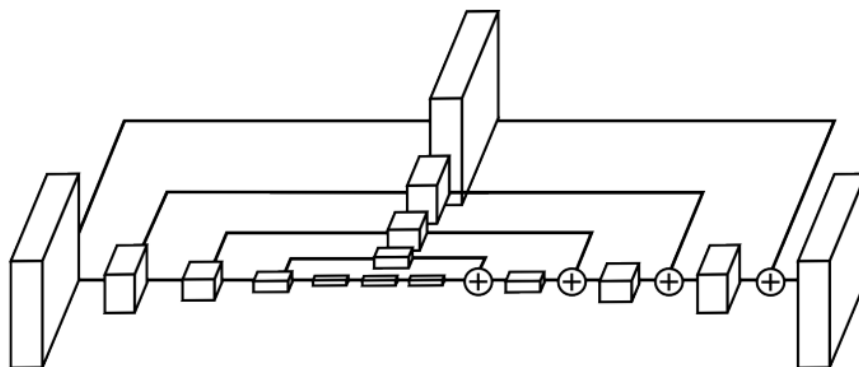


Рис. 1.2 Изображение «hourglass modules»

Для получения предсказаний к крайним блокам применяется свертка 1×1 , которая не изменяет разрешения, но может изменять количество каналов. В качестве предсказаний получаются тепловые карты. Одна карта используется для определения одного сустава.

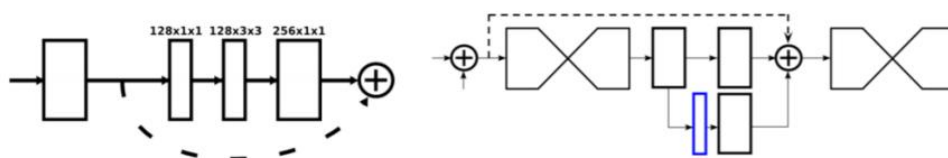


Рис 1.3 Слева: как выглядит «residual module». Справа как выглядит объединение нескольких «hourglass modules», синим помечены предсказанные тепловые карты

Разрешение изображения, которое используется в «hourglass modules» равно 64×64 , когда входное изображение имеет разрешение 256×256 , что также позволяет повысить производительность. Поэтому перед тем как подать изображение в «hourglass» используется свертка 7×7 с шагом 2 и «residual module» с последующей субдискретизацией тем самым достигалась трансформация из 256×256 в 64×64 . «residual module» возвращает тензор с 256 каналами и не меняет разрешение.

1.1.2 PyraNet. Learning Feature Pyramids for Human Pose Estimation

Авторы статьи Learning Feature Pyramids for Human Pose Estimation [4] предлагают архитектуру, которая почти ничем не отличается от Stacked Hourglass Network. Единственное отличие состоит в том, что вместо «residual block» используются PRM (Pyramid Residual Module) блоки. Сеть изображена на рисунке 1.4

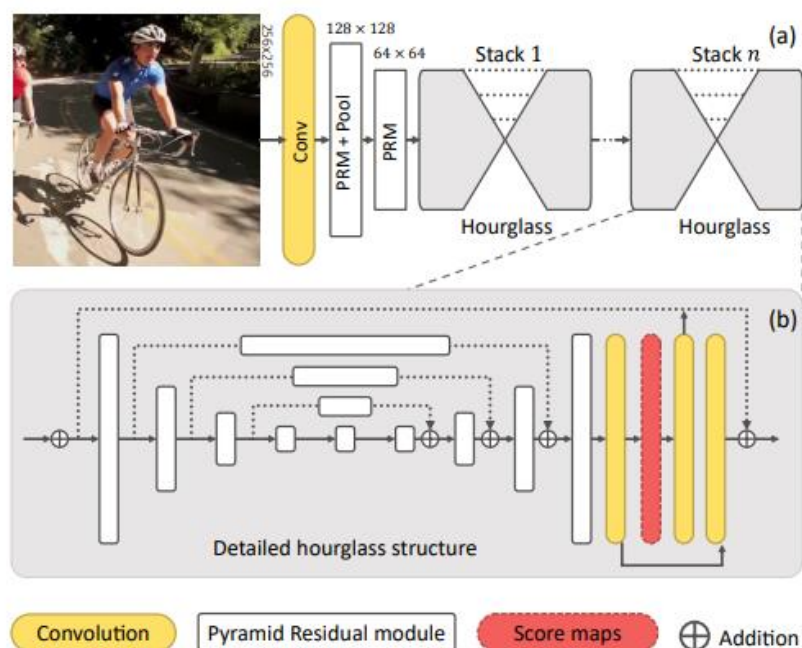


Рис 1.4 Архитектура сети основанной на PRM блоках

Предлагается что «hourglass modules» основанные на PRM- блоках, обрабатывают снимки сделанные в разных масштабах лучше чем «hourglass modules» основанные на «residual modules». Также как видно, каскадирование «hourglass modules» и метод получения тепловых карт не отличается от Stacked Hourglass Network. Рассмотрим как выглядит PRM блоки подробнее. Несколько видов PRM блоков изображены на рисунке 1.5

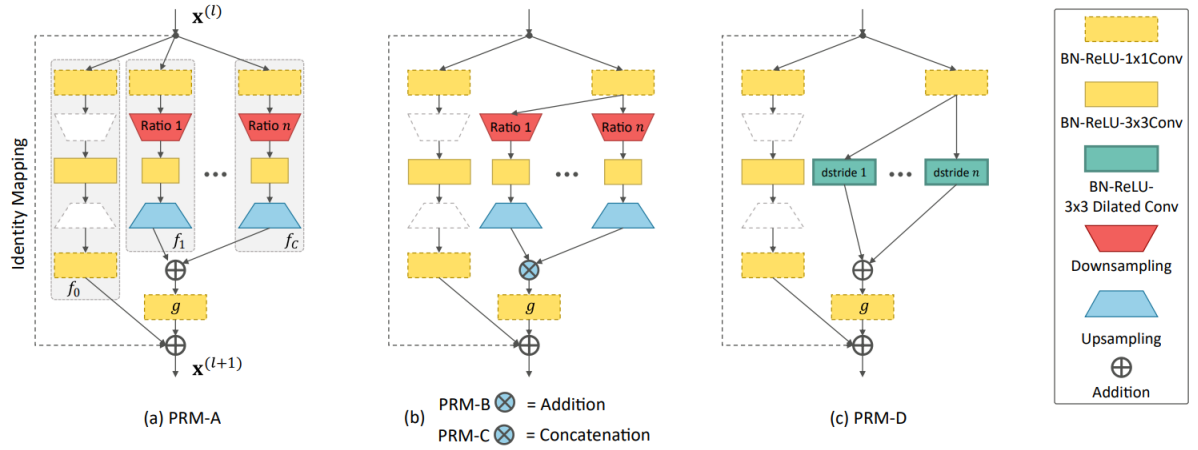


Рис 1.5 PRM блоки

Как видно PRM блоки являются полносверточной сетью и имеют слои изменения дискретизации, что позволяет обрабатывать на каждом этапе трансформации информацию в разных масштабах. Также используется батч-нормализация [7]. Опишем математически технологию данного блока. Пусть у нас входные данные имеют вид $x^{(l)}$ и $W^{(l)}$, тогда выход PRM блока будет равен:

$$x^{(l+1)} = x^{(l)} + \Theta(x^{(l)}; W^{(l)})$$

$\Theta(x^{(l)}; W^{(l)})$ – результат преобразования $x^{(l)}$ с помощью PRM блока

$$\Theta(x^{(l)}; W^{(l)}) = g\left(\sum_{c=1}^C f_c(x^{(l)}; w_{f_c}^{(l)}); w_g^{(l)}\right) + f_0(x^{(l)}; w_{f_0}^{(l)})$$

Где c - это номер ответвления, $W^{(l)} = \{w_{f_c}^{(l)}, w_g^{(l)}\}$ - множество параметров, $f_c(\cdot)$ - функция преобразования $x^{(l)}$, $g(\cdot)$ - функция преобразования суммы или конкатенации $f_c(\cdot)$.

Также авторы в статье указывают, что при таких преобразованиях есть свойство накапливать дисперсию, поэтому рекомендуется для генерации тепловых карт не использовать большие отклонения.

1.2 Обучение нейросети

Для обучения нейросети использовался оптимизатор RMSProp [12] с шагом $lr=0.00002$, в качестве функции ошибки была выбрана MSE (Mean Squared Error). Размер батча был равен 8 изображениям. Сеть обучалась 50 эпох, около 36 часов. В качестве правильных ответов использовалась функция Гаусса [5] с $\sigma_x = 1.5$ и $\sigma_y = 1.5$.

1.2.1 Тепловая карта

Чтобы обучить сеть нам нужны целевые переменные. В самых ранних архитектурах просто предсказывался вектор координат суставов $(x_0, y_0, x_1, y_1, \dots, x_s, y_s)$, где s – количество суставов. То есть авторы рассматривали эту задачу как задачу регрессии. В данном же случае мы используем тепловые карты, чтобы локализовать суставы, затем из этих тепловых карт извлечь координаты. Рассмотрим подробнее как выглядят данные тепловые карты.

Пусть у нас множество истинных координат $\mathbf{z} = \{\mathbf{z}_k\}_{k=1}^K$, где $\mathbf{z}_k = (x_k, y_k)$, обозначает координаты k -го сустава. Тогда тепловая карта \mathbf{S}_k будет сгенерирована с помощью функции Гаусса со средним значением \mathbf{z}_k и отклонением Σ . Следовательно

$$\mathbf{S}_k(\mathbf{p}) \sim \mathcal{N}(\mathbf{z}_k, \Sigma)$$

где $\mathbf{p} \in R^2$ обозначает локацию, Σ – диагональная матрица размером 2×2 , на диагонали которой находятся отклонения по X и Y (σ_x, σ_y). Пример тепловой карты представлен на рисунке. Пусть каждый «hourglass module» предсказывает K -тепловых карт (K -кол-во суставов), то есть $\hat{\mathbf{S}} = \{\hat{\mathbf{S}}_k\}_{k=1}^K$. Тогда функция потерь для каждого «hourglass module» будет выглядеть следующим образом:

$$L = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^K \|\mathbf{s}_k - \hat{\mathbf{s}}_k\|^2$$

Где N - количество изображений в батче.

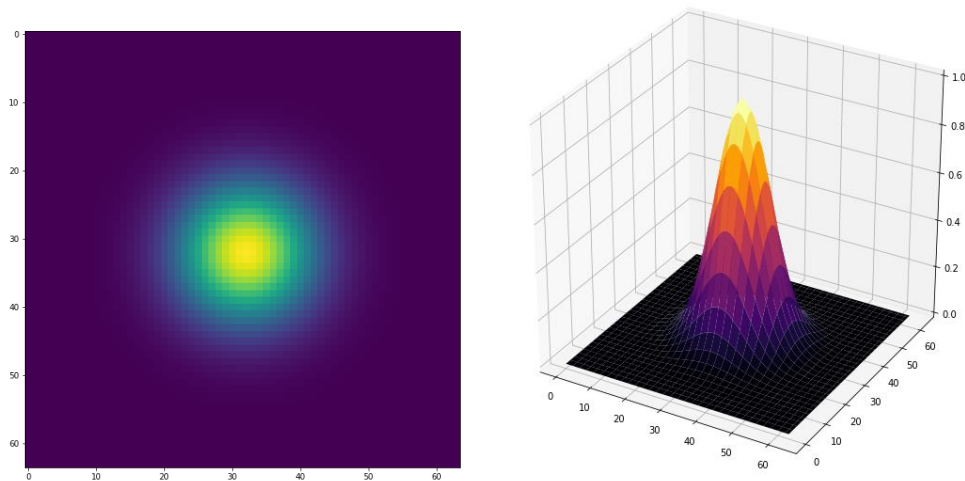


Рис 1.6 Тепловая карта в 3D (справа) и её проекция на плоскость XY (слева)

1.3 Обработка тепловых карт

После предсказания тепловой карты. Нам нужно как-то извлечь из неё координаты. Это можно делать разными способами, рассмотрим 4 из них:

1. Применение `argmax` к тепловой карте
2. Найти «центр тяжести» карты
3. Немного сделать сдвиг от максимального значения ко второму по величине максимальному значению
4. Оценить μ с помощью функции максимального правдоподобия.

1.3.1 Применение `argmax` к тепловой карте

Этот метод является самым простым, в данном случае:

$$\hat{\mathbf{z}}_k = \operatorname{argmax} \mathbf{S}_k(\mathbf{p}), k = 1..K$$

1.3.2 Нахождение «центра тяжести» тепловой карты

В данном случае координаты считаются следующим образом:

$$\hat{\mathbf{z}}_k = (x_k, y_k) = \left(\frac{\sum_i x_i \cdot p_i}{\sum_i p_i}, \frac{\sum_i y_i \cdot p_i}{\sum_i p_i} \right)$$

где p_i – значение тепловой карты в определенном пикселе

1.3.3. Сдвиг ко второму максимальному значению.

Здесь мы вычисляем координаты \mathbf{m} максимального значения и координаты второго максимального значения \mathbf{s} тогда предсказанная координата будет равна:

$$\hat{\mathbf{z}}_k = \mathbf{m} + 0.25 \cdot \frac{\mathbf{s} - \mathbf{m}}{\|\mathbf{s} - \mathbf{m}\|_2}$$

1.3.4. Оценка с помощью функции максимального правдоподобия

Данный способ является самым сложным из всех предыдущих . Авторы статьи Distribution-Aware Coordinate Representation for Human Pose Estimation [6] . Предлагают извлекать координату , учитывая функцию распределения. Ведь максимальное значение функции Гаусса находится в точке μ равному среднему. Пример изображен на рисунке 1.6.

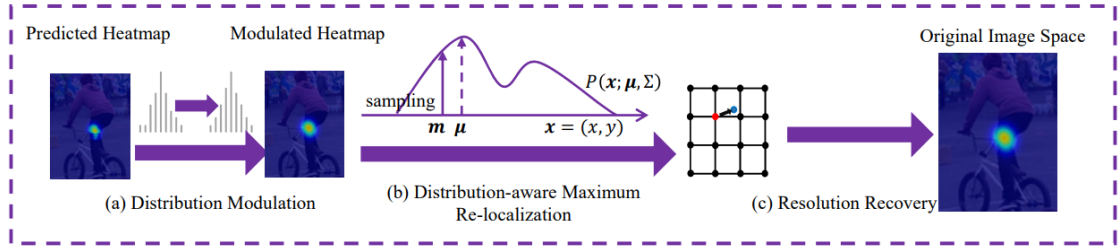


Рис 1.6 Пример из статьи DARK

Рассмотрим данный способ подробнее. Допустим мы предсказали тепловую карту:

$$\Gamma(x; \mu; \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \quad (1)$$

Тогда логарифм функции максимального правдоподобия равен:

$$\Theta(x; \mu; \Sigma) = \ln(\Gamma) = -\ln(2\pi) - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \quad (2)$$

Вычислим первую производную:

$$\mathcal{F}' = \frac{\partial \Theta}{\partial x} = -\Sigma^{-1} (x - \mu) |_{x=\mu} = 0 \quad (3)$$

Разложим $\Theta(\mu)$ в ряд Тейлора в окрестности максимального значения m :

$$\Theta(\mu) = \Theta(m) + \mathcal{F}'(m)(\mu - m) + \frac{1}{2} (\mu - m)^T \mathcal{F}''(m) (\mu - m)$$

Где $\mathcal{F}''(m)$ вторая производная (т.е. Гессиан)

$$\mathcal{F}''(m) = \mathcal{F}''(x)_{x=m} = \Sigma^{-1}$$

Из (5), (4), (3) Получаем:

$$\mu = m - \mathcal{F}''(m)^{-1} \mathcal{F}'(m)$$

Посчитать производные мы можем с помощью формул из численного дифференцирования:

$$f'(x) = \frac{f(x+h)-f(x-h)}{2h} \quad (6)$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{f(x-2h,y)-2f(x,y)+f(x+2h,y)}{(2h)^2} \quad (7)$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{f(x+h,y+h)-f(x+h,y-h)-f(x-h,y+h)+f(x-h,y-h)}{(2h)^2} \quad (8)$$

Чтобы тепловая карта была похожа на Гауссовское распределение стоит применить Гауссовское ядро [9] $\mathbf{h}' = \mathbf{K} \odot \mathbf{h}$, где \odot – операция свертки

2.1 Используемые данные

В данном разделе описываются данные, с помощью которых проходило обучение нейросети, их обработку и используемые аугментации.

2.1.1 Датасет МРП Human Pose

Датасет МРП Human Pose [11] – это один из основных наборов данных, для оценки позы человека. Набор данных включает около 25 тыс. изображений, содержащих более 40 тыс. человек с аннотированными суставами тела. Изображения были извлечены из видео YouTube, каждому изображению присвоена метка с видом деятельности, которым занимается человек на фото. Всего данные охватывают около 410 видов деятельности.

В качестве аннотаций нам даны координаты 16-ти суставов (левый локоть, правое запястье, бедро и т.д.), координата центра ограничивающей рамки, высота ограничивающей рамки деленной на 200.

2.1.2 Аугментация данных

Одним из важных критериев качественного обучения сети, являются хорошо подготовленные данные. Будем использовать аугментации, которые используются в этом справочнике [8]. А именно будем использовать повороты не более чем на 30° , изменять масштаб в пределах от 0.75 до 1.25 и приводить к размеру 256x256. Из-за того, что на фотографии может быть изображено несколько человек или когда человек намного меньше относительно фона. Мы будем вырезать из изображения ту часть фотографии, которая находится в пределах ограничивающей рамки. З Вырезать будем так, чтобы сохранялись пропорции человека и он находится примерно в центре изображения. Рассмотрим более подробно как делать такой вырез.

Вспользуемся матрицей: $\mathbf{A} = \begin{pmatrix} s_x & 0 & t_x \\ 0 & s_y & t_y \\ 0 & 0 & 1 \end{pmatrix}$, где s_x, s_y -изменение масштаба

по осям X и Y, t_x, t_y – смещение по осям X и Y. Пусть размер изображения, который будем подавать нейросети будет равен $r \times r$, а размер ограничивающей рамки равен $h \times h$ и ее центр имеет координаты $\mathbf{c} = (c_x, c_y)$.

Тогда значения $s_x = s_y = \frac{h}{r}$, $t_x = c_x - 0.5 \cdot h$, $t_y = c_y - 0.5 \cdot h$. Найдем левую верхнюю точку и правую нижнюю точку ограничивающей рамки.

Их координаты будут равны $\mathbf{u} = \mathbf{A} \cdot (0,0,1)^T = (c_x - 0.5 \cdot h, c_y - 0.5 \cdot h, 1)^T$,

$\mathbf{b} = \mathbf{A} \cdot (r, r, 1)^T = (c_x + 0.5 \cdot h, c_y + 0.5 \cdot h, 1)^T$, где последнюю единицу мы не учитываем. Но есть одна проблема, в данных очень часто ограничивающая рамка выходит за границы изображения, значит в тех местах где она вышла за пределы будем добавлять нули (padding). Для поворота изображения воспользуемся следующей матрицей:

$$\mathbf{R} = \begin{pmatrix} \alpha & \beta & -\alpha \cdot c_x - \beta \cdot c_y + \frac{r}{2} \\ -\beta & \alpha & -\beta \cdot c_x - \alpha \cdot c_y + \frac{r}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

где $\alpha = \frac{r}{h} \cdot \cos \varphi$, $\beta = \frac{r}{h} \cdot \sin \varphi$, φ -угол поворота. Пример аугментаций изображений представлен на рисунках 2.1 и 2.2



Рис 2.1 Изображение, где ограничивающая рамка не выходит за пределы

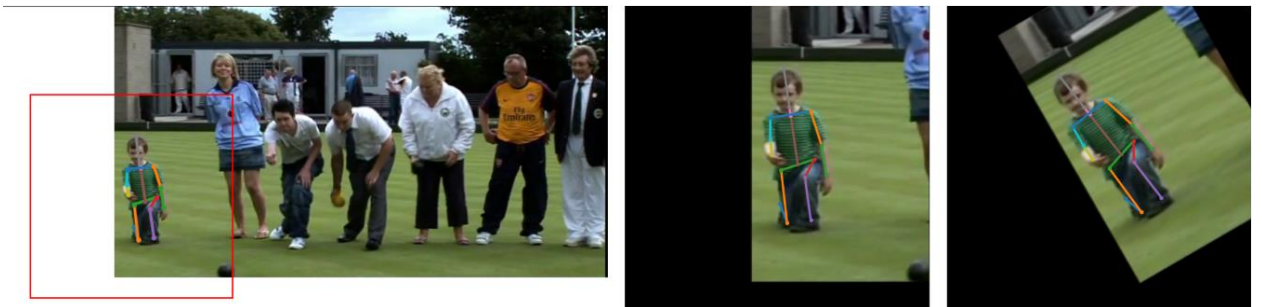


Рис 2.2 Изображение, где ограничивающая рамка выходит за пределы

3.1 Метрика

Качество сети будет определяться по следующей метрике: Процент правильных ключевых точек (PCKh@0.5). В данном случае точка является правильно найденной, если $\rho(\mathbf{z}_k, \hat{\mathbf{z}}_k) \leq 0.5 \cdot \text{headsize}$, где $\rho(\cdot)$ - Евклидово расстояние [9], $\mathbf{z}_k, \hat{\mathbf{z}}_k$ - истинные и предсказанные координаты k -го сустава, $\text{headsize} = \frac{w}{10}$, w - длина или ширина тепловой карты. То есть в нашем случае $\text{headsize} = \frac{64}{10} = 6.4$

4.1 Результаты

Возьмем несколько фотографий из интернета и сделаем предсказания. Примеры предсказаний модели показаны на рисунке 4.1.



Рис 4.1 Примеры предсказаний с помощью нейросети

Можно увидеть, что нейросеть вполне неплохо определяет точки в легких позах, но в сложных позах сеть начинается путаться.

Занесем в таблицу 1 результаты работы сети с разными подходами обработки тепловых карт.

Таблица 1. Сравнение способов обработки тепловых карт по метрике PCKh@0.5 на валидационных данных MPII Human Pose

Метод	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean
PyraNet+(1)	93.6	87.7	73.4	60.0	80.7	63.9	57.9	76.0
PyraNet+(4)	93.9	87.9	73.5	60.8	80.5	63.8	58.7	76.3
PyraNet+(2)	84.8	86.0	66.3	41.3	76.0	45.2	43.6	66.6
PyraNet+(3)	93.5	87.2	72.6	59.3	79.3	62.1	56.8	75.2

Где (1) - Применение argmax к тепловой карте, (2) - Нахождение «центра тяжести» тепловой карты, (3) - Сдвиг ко второму максимальному значению, (4) - Оценка с помощью функции максимального правдоподобия.

Как видно лучшее качество получается при использовании (4)-го способа. Хуже всех оказался (2)-ой способ, связано это с тем, что модель недостаточно обучилась и её ответы бывают не очень «уверенными», пример приведен на рисунке 4.2 . Из-за этого идет сильное смещение в сторону помех, из-за чего качество начинает падать.

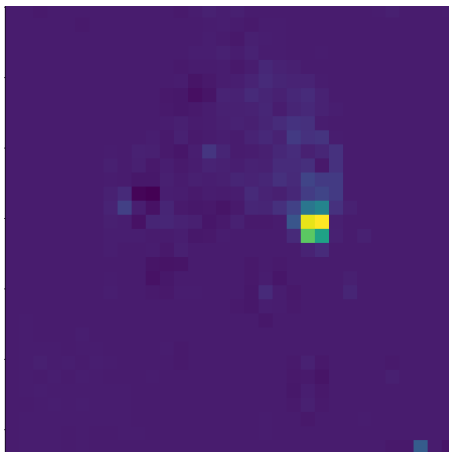


Рис 4.2 Пример предсказанной тепловой карты с помехами

Заключение

В данной работе была изучена релевантная литература, на основе которой была построена и обучена каскадная нейросетевая архитектура, с помощью которой удавалось локализовать суставы человека, также были изучены различные способы извлекать координаты из тепловых карт. Были изучены способы аугментации данных. Дополнительными вопросами, заслуживающего рассмотрения, в данном случае является использование вместо MSE другой функции потерь, например L1, модифицировать (2) способ обработки тепловой карты, попробовать использовать « взвешенный центр тяжести» , добавить шума в данные. Также для тестовых данных предсказывать ключевые точки, тем самым получая примерную ограничивающую рамку вокруг человека и подать в сеть изображение находящиеся внутри рамки и сделать финальное предсказание.

Список литературы

1. Newell, A. Stacked hourglass networks for human pose estimation / A. Newell, K. Yang, J. Deng // European conference on computer vision / Springer. – 2016. – С. 483–499.
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Computer Vision and Pattern Recognition, 2016. CVPR 2016. IEEE Conference on (2015)
3. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. Unet: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.
4. W. Yang, S. Li, W. Ouyang, H. Li, and X. Wang. Learning feature pyramids for human pose estimation. In ICCV, pages 1290–1299, 2017.
5. А.Н. Ширяев. Вероятность. Том 1. МЦНМО 2007, с.330-333.
6. F. Zhang, X. Zhu, H. Dai, M. Ye, and C. Zhu, “Distribution-aware coordinate representation for human pose estimation,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 7093–7102.
7. Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference arxiv, 2022.
8. Aiden Nibali, Zhen He, Stuart Morgan, Luke Prendergast. 2018. ArXiv Prepr. ArXiv 1801.07372
9. Agarap, A.F. Deep learning using rectified linear units (relu) / A.F. Agarap // ArXiv Prepr. ArXiv180308375. – 2018.
10. X. Wan, G. Wang, X. Wei, J. Li, and G. Zhang, “Star centroiding based on fast gaussian fitting for star sensors,” Sensors (Basel, Switzerland), vol. 18, 2018.
11. M. Andriluka, L. Pishchulin, P. Gehler and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3686-3693, doi: 10.1109/CVPR.2014.471.
12. T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning, 2012.

13. Nearing, James (2010). Mathematical Tools for Physics. Chapter 7. Operators and Matrices
14. Shapiro, L. G. & Stockman, G. C. Computer Vision. — Prentice Hall, 2001. — C. 154.