



## ПРЕДСТАВЛЕНИЕ КООРДИНАТ С УЧЕТОМ РАСПРЕДЕЛЕНИЯ ДЛЯ ОЦЕНКИ ПОЗЫ ЧЕЛОВЕКА

Мадаминов Хайдар Худаярович<sup>1</sup>

PhD, доцент,

<sup>1</sup>ТУИТ имени Мухаммада ал-Хоразмий  
Узбекистан, г. Ташкент

Худайберганов Журабек Давлатбоевич<sup>2</sup>

докторант,

<sup>2</sup>ТУИТ имени Мухаммада ал-Хоразмий  
Узбекистан, г. Ташкент

[jurabekkhudayberganov1991@gmail.com](mailto:jurabekkhudayberganov1991@gmail.com)

Каримова Айкерим Отесиновна<sup>3</sup>

ассистент,

<sup>3</sup>НФ ТУИТ имени Мухаммада ал-Хоразмий  
Каракалпакстан, г. Нукус

[karimovaajkerim4@gmail.com](mailto:karimovaajkerim4@gmail.com)

Ешниязова Гоззал Бахтияровна<sup>4</sup>

ассистент,

<sup>3</sup>НФ ТУИТ имени Мухаммада ал-Хоразмий  
Каракалпакстан, г. Нукус

[gozzal1115@gmail.com](mailto:gozzal1115@gmail.com)

<https://doi.org/10.5281/zenodo.7107470>

### ARTICLE INFO

Received: 17<sup>th</sup> September 2022

Accepted: 19<sup>th</sup> September 2022

Online: 23<sup>rd</sup> September 2022

### KEY WORDS

Обширные эксперименты показывают, что DARK дает наилучшие результаты на двух распространенных тестах, MPII и COCO, последовательно подтверждая полезность и эффективность нашей новой идеи представления координат.

### ABSTRACT

В этой научной статье представлено представление координат с особым акцентом на тепловую карту. Интересно, что мы обнаружили, что процесс декодирования предсказанных тепловых карт в окончательные координаты суставов в исходном пространстве изображения на удивление важен для оценки позы человека, что, тем не менее, ранее не признавалось. В свете обнаруженной важности мы дополнительно исследуем конструктивные ограничения стандартного метода декодирования координат, широко используемого существующими методами, и предлагаем более принципиальный метод декодирования с учетом распределения. Обширные эксперименты показывают, что DARK дает наилучшие результаты на двух распространенных тестах, MPII и COCO, последовательно подтверждая полезность и эффективность нашей новой идеи представления координат.

### Введение

Оценка позы человека — фундаментальная проблема

компьютерного зрения, целью которой является определение пространственного положения (т. е.



координаты) суставов человеческого тела на неограниченных изображениях (Андрилука и др., 2014). Это нетривиальная задача, так как внешний вид суставов тела сильно различается из-за разных стилей одежды, произвольной окклюзии и неограниченных фоновых контекстов, в то время как необходимо определить точные координаты суставов. В качестве надежных моделей обработки изображений сверточные нейронные сети (CNN) превосходно справляются с этой задачей (LeCun et al., 1998). Существующие работы обычно сосредоточены на разработке архитектуры CNN, специально предназначенной для вывода о позе человека (Newell, Yang, and Deng, 2016; Sun et al., 2019). По аналогии с обычными однократными векторами в качестве представления метки класса объекта в классификации изображений, CNN-модель позы человека также требует представления метки для кодирования меток координат суставов тела, чтобы можно было количественно определить и вычислить контролируемую потерю обучения во время обучения и совместные координаты могут быть выведены правильно<sup>1</sup>. Де-факто стандартным представлением меток является координатная тепловая карта,

сгенерированная как двумерное распределение/ядро Гаусса с центром в отмеченной координате каждого сустава (Tompson et al. 2014). Он получается в процессе кодирования координат, от координаты до тепловой карты. Тепловая карта характеризуется предоставлением пространственной поддержки вокруг истинного местоположения с учетом не только контекстуальных подсказок, но и присущей неоднозначности положения цели. Важно отметить, что это может эффективно снизить риск переобучения модели при обучении в том же духе, что и регуляризация сглаживания меток классов (Szegedy et al. 2016). Неудивительно, что современные модели поз (Newell, Yang, and Deng, 2016; Xiao, Wu, and Wei, 2018; Sun et al., 2019) основаны на представлении координат тепловой карты. также присущая неоднозначность положения цели. Важно отметить, что это может эффективно снизить риск переобучения модели при обучении в том же духе, что и регуляризация сглаживания меток классов (Szegedy et al. 2016). Неудивительно, что современные модели поз (Newell, Yang, and Deng, 2016; Xiao, Wu, and Wei, 2018; Sun et al., 2019) основаны на представлении координат тепловой карты.

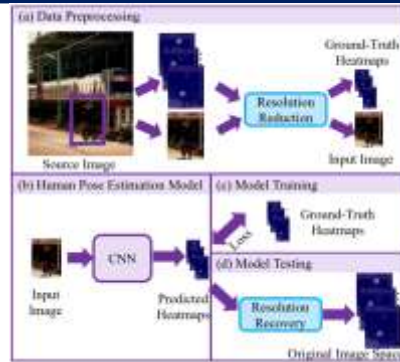


Рисунок 1: Конвейер системы оценки позы человека.

Для повышения эффективности часто применяется уменьшение разрешения к исходным ограничивающим рамкам обнаружения человека, а также к наземному наблюдению тепловой карты. То есть модель работает в пространстве изображений с низким разрешением. Следовательно, во время тестирования необходимо соответствующее восстановление разрешения, чтобы получить совместное предсказание координат в исходном пространстве изображения.

С представлением метки тепловой карты одним из основных препятствий является то, что вычислительные затраты являются квадратичной функцией разрешения входного изображения, что не позволяет моделям CNN обрабатывать необработанные данные изображений с высоким разрешением, как правило. Чтобы быть доступным с вычислительной точки зрения, стандартная стратегия (см. рис. 1) заключается в понижении разрешения всех изображений ограничивающей рамки человека с произвольно большим разрешением до небольшого разрешения с префиксом с помощью процедуры предварительной обработки данных, прежде чем они будут переданы в модель оценки позы человека. Чтобы предсказать

местоположение сустава в исходном координатном пространстве изображения, после предсказания тепловой карты требуется соответствующее восстановление разрешения для преобразования обратно в исходное координатное пространство. Окончательный прогноз рассматривается как место с максимальной активацией. Мы называем этот процесс декодированием координат, от тепловой карты к координате. Стоит отметить, что ошибка квантования может быть внесена при указанном выше уменьшении разрешения. Чтобы облегчить эту проблему, во время существующего процесса декодирования координат обычно выполняется операция сдвига вручную в соответствии с направлением от самой высокой активации к второй по величине активации (Ньюэлл, Ян и Дэн, 2016).

Вклад этой работы заключается в том, что мы обнаруживаем ранее нереализованное значение представления координат в оценке позы человека и предлагаем новый метод представления координат с учетом распределения (DARK) с двумя ключевыми компонентами: (1) эффективное основанное на



разложении Тейлора декодирование координат и (2) несмещенное кодирование координат с центрированием подпикселя. Важно отметить, что существующие методы определения позы человека можно легко использовать с помощью DARK без какой-либо модификации алгоритма. Обширные эксперименты с двумя распространенными эталонными тестами (MPII и COCO) показывают, что наш метод обеспечивает значительное повышение производительности для существующих современных моделей оценки позы человека (Sun et al., 2019; Xiao, Wu, and Wei, 2018; Newell, Yang), and Deng 2016), достигая наилучшей точности одиночной модели в COCO и MPII. DARK выгодно позволяет использовать меньшее разрешение входного изображения с гораздо меньшим снижением производительности, в то же время значительно повышая эффективность логического вывода модели, тем самым облегчая приложения с малой задержкой и низким энергопотреблением, как это требуется во встроженных сценариях искусственного интеллекта.

## **Анализ исследование**

Существует два распространенных дизайна представления координат при оценке позы человека: прямая координата и тепловая карта. Оба используются в качестве целей регрессии для обучения модели. Координатная регрессия Непосредственное использование координат в качестве выходных данных модели является простым и интуитивно понятным. Но только несколько существующих методов используют

этот дизайн (Тошев и Сегеди, 2014; Фан и др., 2015; Каррейра и др., 2016; Сан и др., 2018). Одна из вероятных причин заключается в том, что в этом представлении отсутствует пространственная и контекстуальная информация, что делает изучение модели позы человека чрезвычайно сложным из-за внутренней визуальной неоднозначности в совместном местоположении. Регрессия тепловой карты Представление тепловой карты элегантно устраняет вышеуказанные ограничения. Впервые он был представлен в (Tompson et al. 2014) и быстро стал наиболее часто используемым представлением координат. Как правило, основные исследования сосредоточены на разработке сетевых архитектур для более эффективного регресса наблюдения за тепловыми картами. Репрезентативные усовершенствования дизайна включают последовательное моделирование (Gkioxari, Toshev, and Jaitly, 2016; Belagiannis and Zisserman, 2017), расширение рецептивного поля (Wei et al., 2016), позиционное голосование (Lifshitz, Fetaya, and Ullman, 2016), промежуточный надзор (Newell, Yang), and Deng 2016; Wei et al. 2016), моделирование парных отношений (Chen and Yuille 2014), моделирование древовидной структуры (Chu et al. 2016b; Yang et al. 2016; Chu et al. 2016a; Sun et al. 2017; Tang, Yu, and Wu, 2018), пирамидальное остаточное обучение (Yang et al., 2017), каскадное пирамидальное обучение (Chen et al., 2018), обучение, управляемое знаниями (Ning, Zhang, and He, 2017), активное обучение (Liu и Ferrari, 2017), состязательное обучение (Chen et al.,



2017), повышение дискретизации с деконволюцией (Xiao, Wu, and Wei, 2018), многомасштабное наблюдение (Ke et al., 2018), механизм внимания (Liu et al., 2018; Su et al. al. 2019) и сохранение представления с высоким разрешением (Sun et al. 2019). В отличие от всех предыдущих работ, мы вместо этого исследуем вопросы представления тепловой карты при оценке позы человека, что в значительной степени игнорируется в литературе. Мы не только выявляем большое влияние снижения разрешения в процессе использования тепловой карты, но и предлагаем принципиальный метод представления координат для значительного улучшения производительности существующих моделей. Важно отметить, что наш метод можно легко интегрировать без изменения конструкции модели.

### *Методология*

Мы рассматриваем проблему представления координат, включая кодирование и декодирование при оценке позы человека. Цель состоит в том, чтобы предсказать совместные координаты в заданном входном изображении. С этой целью нам необходимо изучить модель регрессии от входного изображения к выходным координатам, а тепловая карта часто используется как представление координат как во время обучения модели, так и во время тестирования. В частности, мы предполагаем доступ к обучающему набору изображений. Чтобы облегчить обучение модели, мы кодируем помеченную наземную координату сустава в тепловую карту в качестве контролируемой цели

обучения. Затем во время тестирования нам нужно декодировать предсказанную тепловую карту в координату в исходном координатном пространстве изображения. Далее мы сначала опишем процесс декодирования, сосредоточив внимание на анализе ограничений существующего стандартного метода и разработке нового решения. Затем мы дополнительно обсудим и рассмотрим ограничения процесса кодирования. Наконец, мы описываем интеграцию существующих моделей оценки позы человека с предлагаемым методом.

### *Расшифровка координат*

Несмотря на то, что это считается незначительным компонентом конвейера тестирования моделей, как мы обнаружили в этом исследовании, декодирование координат оказывается одним из наиболее значительных факторов, влияющих на производительность при оценке позы человека на изображениях (см. Таблицу 1). В частности, это процесс перевода предсказанной тепловой карты каждого отдельного сустава в координату в исходном пространстве изображения. Предположим, что тепловая карта имеет тот же пространственный размер, что и исходное изображение, нам нужно только найти место максимальной активации в качестве совместного предсказания координат, что является прямым и простым. Однако это часто не так, как интерпретировано выше. Вместо этого нам нужно увеличить разрешение тепловых карт до исходного разрешения изображения с помощью неограниченного фактора, зависящего от выборки,  $\lambda \in \mathbb{R}^+$ . Это связано с проблемой субпиксельной

локализации. Прежде чем представить наш метод, мы сначала вернемся к стандартному методу декодирования координат, используемому в существующих моделях оценки позы. Стандартный метод декодирования координат разработан эмпирически в соответствии с характеристиками модели (Newell, Yang, and Deng, 2016). В частности, учитывая тепловую карту  $h$ , предсказанную обученной моделью, мы сначала определяем координаты максимальной ( $m$ ) и второй максимальной ( $s$ ) активации. Затем совместное местоположение прогнозируется как:

$$p = m + 0.25 \frac{8 - m}{\|8 - m\|_2}$$

где  $\|8 - m\|_2$  определяет величину вектора. Это означает, что прогноз представляет собой максимальную активацию со смещением на 0,25 пикселя (т. е. субпикселя) в сторону второй максимальной активации в пространстве тепловой карты. Окончательное предсказание координат в исходном изображении вычисляется как:

$$\hat{p} = \lambda p$$

где  $\lambda$  — коэффициент уменьшения разрешения.

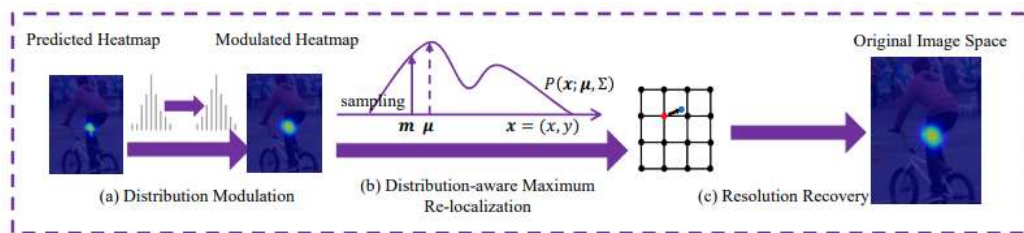
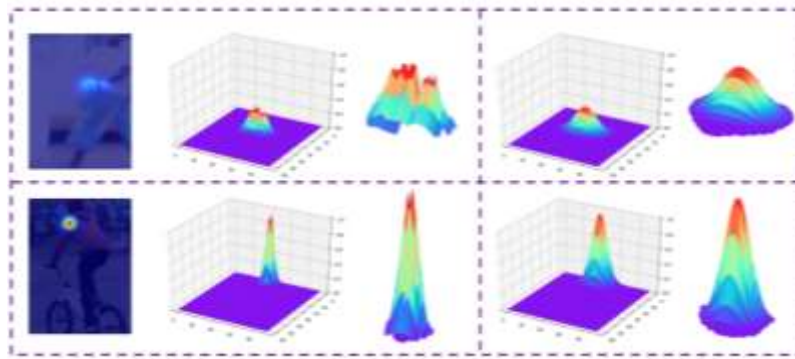


Рисунок 2: Обзор предлагаемого метода декодирования координат с учетом распределения.

Модуляция распределения тепловой карты поскольку предлагаемый метод декодирования координат основан на предположении о Гауссовском распределении, нам необходимо проверить, насколько хорошо выполняется это условие. Мы обнаружили, что тепловая карта часто модулирует распределение тепловой карты. Поскольку предлагаемый метод декодирования координат основан на предположении о распределении по Гауссу, нам необходимо проверить, насколько хорошо выполняется это условие. Мы обнаружили, что часто

тепловая карта, предсказанная моделью оценки позы человека, не демонстрирует гауссовой структуры хорошей формы по сравнению с данными обучающей тепловой карты. Как показано на рис. 3(a), тепловая карта обычно представляет несколько пиков вокруг максимальной активации. Это может отрицательно сказаться на производительности нашего метода декодирования. Чтобы решить эту проблему, мы предлагаем заранее модулировать распределение тепловой карты.



(a) Прогнозируемая тепловая карта (b) Модулированная тепловая карта

Рисунок 3: Иллюстрация модуляции распределения тепловой карты. (a) прогнозируемая тепловая карта; (b) Модулированное распределение тепловой карты.

2. В частности, последовательность включает три этапа: (a) Модуляция распределения тепловой карты.

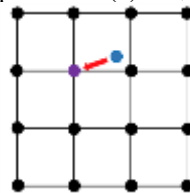


Рисунок 4: Иллюстрация ошибки квантования в стандартном процессе кодирования координат.

Синяя точка обозначает точное положение ( $g_0$ ) сустава. При квантовании координат на основе пола вносится ошибка (указана красной стрелкой). Другие методы квантования имеют ту же проблему.

*Эксперименты наборы данных.* Мы использовали два популярных набора данных для оценки позы человека, COCO и MPII. Набор данных ключевых точек COCO (Lin et al. 2014) представляет собой естественные сложные данные изображений с различными позами человека, неограниченными средами, различными масштабами тела и паттернами окклюзии. Вся цель включает в себя как обнаружение экземпляров человека, так и локализацию суставов тела. Он содержит 200 000 изображений и 250 000 образцов людей. Каждый экземпляр человека помечен 17 суставами. Аннотации наборов для обучения и проверки общедоступны. При оценке мы следовали обычно используемому

разделению train2017/val2017/test-dev2017. Набор данных о позах человека MPII (Andriluka et al. 2014) содержит образцы 40 000 человек, каждый из которых помечен 16 суставами. Мы следовали стандартному разделению поезд/оценка/тест, как в (Tompson et al. 2014).

*Метрики оценки.* Мы использовали сходство ключевых точек объекта (OKS) для COCO и процент правильных ключевых точек (PCK) для MPII для оценки производительности модели.

*Детали реализации* Для обучения модели мы использовали оптимизатор Adam. Для HRNet (Sun et al. 2019) и SimpleBaseline (Xiao, Wu и Wei 2018) мы следовали тому же графику обучения и эпохам, что и в оригинальных работах. Для «Песочных часов» (Newell, Yang, and Deng, 2016) базовая скорость обучения была настроена на  $2,5e-4$  и снизилась до  $2,5e-5$  и  $2,5e-6$  на 90-й и 120-й эпохах. Общее количество эпох равно 140. В наших экспериментах мы использовали



три разных размера входных данных (128 × 96, 256 × 192, 384 × 288). Мы использовали ту же предварительную обработку данных, что и в (Sun et al. 2019).

*Оценка представления координат.* В качестве основной проблемы в этой работе сначала было изучено влияние представления координат на

производительность модели в связи с разрешением (размером) входного изображения. В этом тесте по умолчанию мы использовали HRNet-W32 (Sun et al. 2019) в качестве базовой модели и 128 × 96 в качестве размера входных данных, а также сообщали о результатах точности на проверочном наборе COCO.

Таблица 2

Влияние модуляции распределения (DM) на набор значений COCO. Магистраль: HRNet-W32; Входной размер: 128×96.

DM	AP	AP <sup>50</sup>	AP <sup>75</sup>	AP <sup>M</sup>	AP <sup>L</sup>	AR
Нет	68.1	88.5	77.1	65.8	73.7	74.8
Да	68.4	88.6	77.4	66.0	74.0	74.9

Таблица 3:

Влияние кодирования координат на набор проверки COCO. Модель: HRNet-W32; Размер ввода: 128 × 96.

Encode	Decode	AP	AP <sup>50</sup>	AP <sup>75</sup>	AP <sup>M</sup>	AP <sup>L</sup>	AR
Biased	Standart	66.9	88.7	76.3	64.6	72.3	73.7
Unbiased	Standart	68.0	88.9	77.0	65.4	73.7	74.5
Biased	Ours	68.4	88.6	77.4	66.0	74.0	74.9
Unbiased	Ours	70.7	88.9	78.4	67.9	76.6	76.7

(a) Расшифровка координат. Мы оценили влияние декодирования координат, в частности операции сдвига и модуляции распределения. Использовались обычные тепловые карты со смещением. В этом тесте мы сравнили предложенный метод сдвига с учетом распределения без сдвига (т. е. непосредственно с использованием максимального местоположения активации) и стандартный сдвиг (уравнение (1)). В Таблице 1 мы делаем два основных наблюдения: (i) Стандартное переключение дает повышение точности AP на целых 5,7%, что на удивление эффективно. Насколько нам известно, это первый отчет об анализе эффективности в литературе, поскольку эта проблема в значительной степени игнорируется предыдущими исследованиями. Это раскрывает ранее невиданное значение декодирования координат для оценки

позы человека. (ii) Несмотря на большой выигрыш от стандартного метода декодирования, предлагаемая модель дополнительно улучшает показатель AP на 1,5%, среди которых модуляция распределения дает 0,3%, как показано в таблице 2. Это подтверждает превосходство нашего метода декодирования.

(b) Координатное кодирование. Мы проверили, насколько эффективным может быть кодирование координат. Мы сравнили предлагаемое несмещенное кодирование со стандартным смещенным кодированием, а также со стандартным и нашим методом декодирования. Из Таблицы 3 мы заметили, что наше несмещенное кодирование с точным центром ядра дает положительный запас производительности, независимо от метода декодирования координат. В частности, беспристрастное

кодирование в обоих случаях стабильно

дает более 1% прироста AP.



Рисунок 5: Качественная оценка DARK (красный) по сравнению с HRNetW32 (голубой) на COCO.

Это говорит о важности кодирования координат, которым опять же пренебрегали предыдущие исследования.

(с) Входное разрешение. Мы изучили влияние разрешения/размера входного изображения, протестировав несколько разных размеров, считая, что это важный фактор, относящийся к эффективности вывода модели. Мы сравнили нашу модель DARK (HRNet-W32 в качестве магистрали) с исходной HRNet-W32, используя предвзятую тепловую карту для обучения и стандартное смещение для тестирования. Из таблицы 4 у нас есть несколько наблюдений: (а) при уменьшении размера входного изображения, как и ожидалось, производительность модели постоянно ухудшается, в то время как стоимость логического вывода явно падает. (б) При поддержке DARK можно эффективно снизить потери производительности модели, особенно в случае очень малого входного разрешения (т. е. очень быстрого вывода модели). Это облегчает развертывание моделей оценки позы человека на устройствах с низким уровнем ресурсов, что очень

желательно для появляющегося встроенного ИИ.

(d) Общее. Помимо самой современной HRNet, мы также протестировали две другие репрезентативные модели оценки позы человека в различных архитектурах CNN: SimpleBaseline (Xiao, Wu, and Wei, 2018) и Hourglass (Newell, Yang, and Deng, 2016). Результаты в таблице 5 показывают, что DARK в большинстве случаев обеспечивает значительный прирост производительности по сравнению с существующими моделями. Это говорит об общей полезности нашего подхода. Мы показали качественную оценку на рис. 5.

(е) Сложность. Мы проверили влияние нашего метода на эффективность вывода в HRNet-W32 при размере входных данных  $128 \times 96$ . На графическом процессоре Titan V скорость работы снижается с 360 кадров в секунду до 320 кадров в секунду в низкоэффективной среде Python, т.е. 11%. Мы считаем, что эта дополнительная стоимость вполне доступна.

## Вывод

В этой работе мы впервые систематически исследовали широко игнорируемую, но важную проблему



представления координат (включая кодирование и декодирование) для оценки позы человека на изображениях без ограничений. Мы не только раскрыли подлинную значимость этой проблемы, но также представили новое представление координат ключевой точки с учетом распределения (DARK) для более разборчивого обучения модели и вывода. Выступая в качестве готового к использованию подключаемого компонента, существующие современные модели могут быть легко использованы с помощью нашего метода DARK без

какой-либо алгоритмической адаптации при незначительных затратах. Помимо эмпирической демонстрации важности представления координат, мы подтвердили преимущества DARK в производительности, проведя обширные эксперименты с широким спектром современных моделей на двух сложных наборах данных. Мы также предоставили последовательность углубленного анализа компонентов, чтобы дать представление о конструктивном обосновании формулировки нашей модели.

## References:

1. Begmatov Sh.A., Arabboyev M.M, Xudayberganov J.D. “ Google media-pipe kutubxonasi dan foydalangan holda inson gavdasi harakatlari farqlarini baholash” Muhammad al-Xorazmiy avlodlari 3(21)2022.
2. Andriluka, M.; Pishchulin, L.; Gehler, P.; and Schiele, B. 2014. 2d human pose estimation: New benchmark and state of the art analysis. In IEEE Conference on Computer Vision and Pattern Recognition. [Andriluka et al. 2014]
3. Belagiannis, V., and Zisserman, A. 2017. Recurrent human pose estimation. In IEEE Conference on Automatic Face and Gesture Recognition. [Belagiannis and Zisserman 2017]
4. Carreira, J.; Agrawal, P.; Fragkiadaki, K.; and Malik, J. 2016. Human pose estimation with iterative error feedback. In IEEE Conference on Computer Vision and Pattern Recognition. [Carreira et al. 2016]
5. Chen, X., and Yuille, A. L. 2014. Articulated pose estimation by a graphical model with image dependent pairwise relations. In Advances in Neural Information Processing Systems. [Chen and Yuille 2014]
6. Chen, Y.; Shen, C.; Wei, X.-S.; Liu, L.; and Yang, J. 2017. Adversarial posenet: A structure-aware convolutional network for human pose estimation. In IEEE International Conference on Computer Vision. [Chen et al. 2017]
7. Chen, Y.; Wang, Z.; Peng, Y.; Zhang, Z.; Yu, G.; and Sun, J. 2018. Cascaded pyramid network for multi-person pose estimation. In IEEE Conference on Computer Vision and Pattern Recognition. [Chen et al. 2018]
8. Chu, X.; Ouyang, W.; Li, H.; and Wang, X. 2016a. Structured feature learning for pose estimation. In IEEE Conference on Computer Vision and Pattern Recognition. [Chu et al. 2016a]



9. Chu, X.; Ouyang, W.; Wang, X.; et al. 2016b. Crf-cnn: Modeling structured information in human pose estimation. In *Advances in Neural Information Processing Systems*, 316– 324. [Chu et al. 2016b]
10. Fan, X.; Zheng, K.; Lin, Y.; and Wang, S. 2015. Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*. [Fan et al. 2015]
11. Fang, H.-S.; Xie, S.; Tai, Y.-W.; and Lu, C. 2017. Rmpe: Regional multi-person pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2334–2343. [Fang et al. 2017]
12. Gkioxari, G.; Toshev, A.; and Jaitly, N. 2016. Chained predictions using convolutional neural networks. In *European Conference on Computer Vision*. [Goodfellow, Bengio, and Courville 2016] Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep learning*. MIT press. [Gkioxari, Toshev, and Jaitly 2016]
13. Ke, L.; Chang, M.-C.; Qi, H.; and Lyu, S. 2018. Multi-scale structure-aware network for human pose estimation. In *European Conference on Computer Vision*. [Ke et al. 2018]
14. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P.; et al. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324. [LeCun et al. 1998]