

Министерство образования и науки Российской Федерации
Московский физико-технический институт (национальный
исследовательский университет)

Физтех-школа радиотехники и компьютерных технологий
Кафедра интеллектуальных информационных систем и технологий

Выпускная квалификационная работа магистра

Исследование методов доменной адаптации для
улучшения распознавания ключевых точек на теле
человека

Автор:

Студент М01-205а группы
Токарев Андрей Сергеевич

Научный руководитель:

Доктор технических наук
Назаров Алексей Николаевич

Научный консультант:

Ст. Преподаватель
Воронков Илья Михайлович



Москва 2024

Аннотация

Исследование методов доменной адаптации для
улучшения распознавания ключевых точек на теле
человека

Токарев Андрей Сергеевич

Краткое описание задачи и основных результатов,
мотивирующее прочитать весь текст

Содержание

1	Введение	4
2	Сверточные нейронные сети	8
3	Распознавание ключевых точек	9
3.1	Ключевые точки	9
3.2	Распознавание ключевых точек на теле человека. Скелет человека	11
4	Domain Adaptation	15
4.1	Перенос знаний	15
4.2	Основные определения	17
4.3	Классификация методов доменной адаптации	20
5	Обзор существующих решений	24
5.1	Обзор моделей для распознавания КТ	24
5.2	Анализ и выбор методов доменной адаптации	25
6	Эксперимент	27
6.1	Описание эксперимента	27
6.2	Данные	29
6.3	Результаты эксперимента	30
7	Заключение	31
	Список литературы	32

1 Введение

Современные технологии машинного обучения и компьютерного зрения продолжают активно развиваться, находя применение в самых разнообразных областях. Одно из направлений, активно развивающихся в последние годы, является решение задачи распознавания ключевых точек на теле человека (Keypoint Detection) или оценка позы человека (Human Pose Estimation). В настоящее время методы решения этой задачи могут иметь разнообразные практические применения.

Одной из возможностей использовать распознавание позы человека является виртуальная реальность. Оцифровка позы человека с помощью неросетей позволяет сэкономить на закупке дорогостоящих костюмов. Можно установить несколько камер, которые будут восстанавливать позу человека и переносить ее компьютерное пространство. При добавлении генеративных алгоритмов можно создавать всевозможные аватары и погрузиться в "Оазис" из фильма Стивена Спилберга "Первому игроку приготовиться".

Другим, уже вполне реальным, применением данной технологии является использование её в качестве рефери на спортивных соревнованиях. Уже сейчас система полуавтоматического определения офсайда активно помогает судьям футбольных матчей по всему миру. Эта система функционирует на основе распознавания ключевых точек частей тела футболистов, которыми они могут играть в мяч, что позволяет определить, были ли нарушены правила или гол был забит чисто [1]. Таким образом, технология существенно повышает точность и объективность судейства, уменьшая количество ошибок и спорных моментов в игре. (ССЫЛКА НА ЦИФРЫ ПРИ НАЛИЧИИ)

Продолжая тему спорта, следует отметить, что оценка позы может быть использована для анализа тренировок и создания персональных по-

мощников. Уже существуют несколько решений, направленных на анализ вашей игры в большой теннис, которые способны оценивать текущие результаты, указывать на области, требующие улучшения, и предлагать рекомендации по коррекции техники [2, 3]. Существует также проект MediaPipe от Google, предоставляющий публичные интерфейсы для анализа спортивной активности на основе распознавания ключевых точек [4]. Этот проект не ограничивается только оценкой и классификацией асан йоги. Он также включает функции для подсчёта количества повторяющихся упражнений, таких как отжимания, подтягивания и приседания.

Применений данной технологии можно придумать множество, но для их реализации необходима модель, которая будет работать быстро, поддерживая режим реального времени, а также демонстрировать высокие показатели точности своей работы. Однако обучение модели и разработка алгоритма её работы представляют собой чрезвычайно сложный и трудоемкий процесс. Этот процесс требует значительных ресурсов, как со стороны специалистов, так и в плане вычислительной мощности. Сначала необходимо собрать и подготовить данные, затем обучить модель, настроить её параметры и протестировать на различных наборах данных, чтобы убедиться в её точности и надёжности. Это часто занимает много времени и требует значительных финансовых вложений.

В тоже время, датасеты, на которых обучаются модели часто имеют общий характер и могут вносить сильную погрешность в результаты при изменении общих характеристик входных данных. Например, модель, обученная распознавать объекты на дневных фотографиях, может показывать низкую точность на ночных снимках из-за разницы в освещении и визуальных характеристиках. Поэтому важно проводить дополнительное обучение модели на данных, соответствующих конкретной задаче.

Этот процесс также требует значительных усилий, аналогичных созданию универсального решения. В связи с этим ученые задумались над тем, как можно уменьшить объем дополнительных работ по приспособлению модели к новым проблемам, не теряя в ее производительности. Таким образом появились алгоритмы доменной адаптации (англ. domain adaptation) и переноса обучения (англ. transfer learning).

Суть данных подходов состоит в том, чтобы преодолеть разрыв между исходным и целевым доменами данных. Это достигается путем выравнивания распределений данных, адаптации признаков и применения обученных моделей к новому домену с минимальной дообработкой. Таким образом, модели становятся более гибкими и способны эффективно работать в различных условиях, не требуя значительного объема новых данных или переработки архитектуры. В конечном итоге это не только ускоряет процесс внедрения, но и снижает затраты на разработку, поскольку уменьшает необходимость в полном цикле обучения новой модели.

Оценив все вышеизложенное, можно сделать вывод, что методы доменной адаптации для моделей распознавания ключевых точек на теле человека имеют значительный потенциал. Различные приложения требуют адаптации к специфическим условиям съемки, будь то освещение, ракурс или качество изображения. А в разнообразных видах спорта, акцент может быть смещен на разные части тела: в футболе важны ноги и корпус, а в баскетболе — руки и верхняя часть туловища.

Стоит отметить, что возможность адаптировать нейронную сеть к целевому набору данных имеет особую важность в задаче распознавания ключевых точек на теле человека. Различные приложения требуют адаптации к специфическим условиям съемки, будь то освещение, ракурс или

качество изображения. А в разнообразных видах спорта, акцент может быть смещен на разные части тела: в футболе важны ноги и корпус, а в баскетболе — руки и верхняя часть туловища. Отсюда требуется возможность быстрого и дешевого улучшения качества работы нейронной сети.

С учетом вышесказанного, напрашивается вывод, что тема данной работы является полезной и важной в нынешних реалиях. В разд. 5 произведен обзор как различных моделей распознавания ключевых точек на теле человека (см в разд. 5.1), так и некоторых методов доменной адаптации (см в разд. 5.2). Также проведен эксперимент по применению *Progressive unsupervised learning (PUL)*, который хорошо себя показал в задачах детекции объектов и повторной реидентификации, к оценке позы. Для него был собран и размечен целевой набор данных, описанный в разд. 6. Результаты эксперимента дают ход дальнейшим исследованиям применения PUL, за счет вариативности способов отбора псевдо-разметки.

ПОПРАВИТЬ ОКОНЧАНИЕ. КАК БУДТО НЕКРАСИВОЕ

2 Сверточные нейронные сети

Добавить этот раздел при наличии времени и только при необходимости
увеличения объема работы.

3 Распознавание ключевых точек

С развитием технологий человечество начало ставить все более разнообразные задачи, для решения которых применялись сверточные нейронные сети (convolutional neural network, CNN). Одной из таких задач оказалось распознавание ключевых точек (Keypoint Detection). При этом распознавание ключевых точек на теле человека выделилось в отдельный раздел, известный как оценка позы (Pose Estimation). Далее рассмотрим формулировку этих задач.

3.1 Ключевые точки

Задача распознавания ключевых точек заключается в том, чтобы обнаружить и точно локализовать определенные точки или места внутри изображения или кадра видео. Эти точки, называемые ключевыми, могут быть определены для различных объектов, таких как лица, тела человека или других структур. Например, в случае распознавания лица ключевые точки могут включать углы глаз, кончик носа, уголки рта и другие характерные особенности лица. В контексте человеческого тела ключевыми точками могут быть суставы, такие как локти, колени, плечи и так далее. Для других объектов ими могут быть выбраны уникальные элементы, которые помогают идентифицировать и анализировать заданный предмет.

Другими словами, ключевые точки являются структурными, которые используются для определения положения и/или местонахождения объекта в пространстве. Они играют важную роль в задачах компьютерного зрения, таких как отслеживание движений, 3D-моделирование, анимация, медицинская визуализация и другие. Они могут использоваться для создания каркасов объектов, анализа их формы, измерения расстояний между различными частями и выполнения других аналитических

задач.

Более точно эти объекты можно определить следующим образом: *ключевые точки (КТ)* - это специфические, заранее определенные части распознаваемого объекта, которые имеют особое значение для дальнейшего анализа местоположения объекта на изображении. Каждая ключевая точка обычно соответствует определенной анатомической или структурной особенности, которая легко распознается и может служить ориентиром для алгоритмов обработки изображений. Иными словами, КТ необходимо обладать следующими характеристиками, чтобы можно было использовать их в качестве референсных для заданного объекта:

1. *Уникальность*

Точки должны быть уникальными и отличаться от других точек на изображении

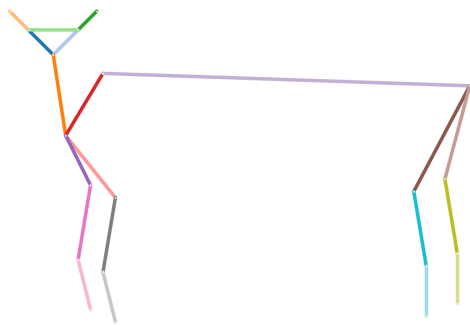
2. *Инвариантность*

Точки должны сохранять свою идентичность при общих преобразованиях изображения, таких как вращение, масштабирование и изменения условий освещения

3. *Повторяемость*

Точки должны быть обнаруживаемы в разных экземплярах одного и того же объекта или сцены

Для анализа структуры объекта и взаимосвязи между его различными КТ часто применяется схематичное описание, которое обеспечивает более наглядное визуальное представление. Этот метод помогает лучше понять анатомическую и функциональную структуру распознаваемого предмета. При этом саму схему, представляющую собой своего рода «скелет», принято называть *топологией*. Примеры различных топологий представлены на рис. 1.



(a) Animal Keypoints



(b) Car keypoints

Рис. 1: Примеры топологий объектов от OpenPifPaf [5]

Математическая постановка задачи

СДЕЛАТЬ ПРИ НАЛИЧИИ ВРЕМЕНИ

3.2 Распознавание ключевых точек на теле человека. Скелет человека

На теле человека тоже можно выделить несколько ключевых точек, информация о которых дает возможность цифровизовать позу человека и использовать ее для аналитики с помощью методов машинного обучения. Именно для этого и была разработана задача распознавания ключевых точек на теле человека или, как ее часто называют в англоязычной литературе, задача оценки позы человека (Human Pose Estimation).

Основной вопросом для НРЕ стал выбор набора ключевых точек и топологии, по которой они будут соединяться. В первых работах были представлены различные примеры топологий, некоторые примеры представлены на рис. 2. Если точки туловища имеют большое количество пересечений, то точки головы сильно отличались: где-то учитывалось только положение головы (то есть добавлена верхняя точка головы), где-то рассматривались некоторые точки на лице, где-то голова вообще не учи-

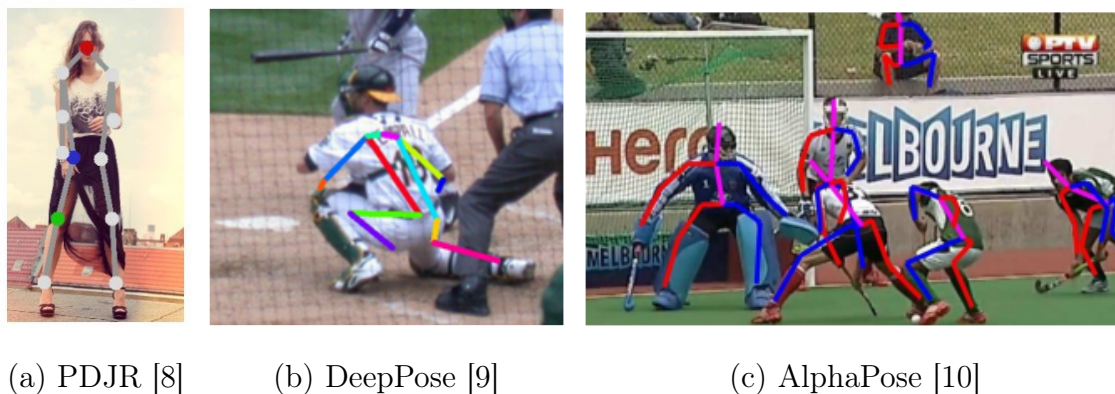


Рис. 2: Примеры различных топологий у первых решений задачи распознавания ключевых точек на теле человека

тывалась. Все зависело от задачи и возможностей исследователей. Позже в 2015 году Microsoft выпустило набор данных с детальным описанием 17 точек на теле человека и запустила соревнование по распознаванию этих точек [6, 7]. Исследователей это заинтересовало и они начали адаптировать свои модели под топологию, описанную в датасете COCO [6]. Хотя данные в наборах перестали обновлять после 2017 года, многие новые модели до сих пор оценивают свои модели по набору данных COCO. Отсюда и получилось, что данная топология стала основной для задачи оценки позы.

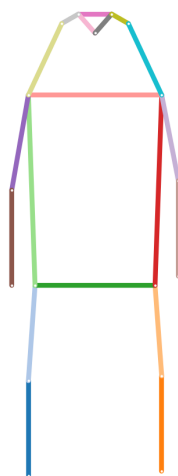


Рис. 3: Топология COCO

Другим вопросом для описываемой задачи стал подход к распознаванию. Делать детекцию человека и уже на кропнутом изображении производить поиск или искать все возможные точки, а потом собирать их в скелет. Эти две идеи и сформировали два направления развития методов НРЕ. Немного о них:

1. *Подход сверху-вниз (англ. top-down)*

Для данного этапа вам понадобится дополнительная модель детекции, которая локализует человека на изображении, выделяя его в прямоугольную область. Затем этот прямоугольник передается в модель распознавания ключевых точек. Этот подход обеспечивает высокую точность предсказания КТ, но может быть чувствителен к ошибкам на этапе обнаружения и ориентации человека в кадре, что требует дополнительной предобработки изображений.

2. *Подход внизу-вверх (англ. bottom-up)*

Для данного подхода вам не нужны помощники в виде детектора, но все равно имеет два этапа в своей работе. Первоначально модель распознает все ключевые точки на полученном на вход изображении, получая таким образом карту распределения КТ по фото. А вторым шагом модель собирает все полученные точки единый скелет. При детекции нескольких людей необходимо верно сопоставить их части тела. Для этого есть несколько способов, один из которых является построение полей сходства частей тела (part affinity fields), используемый в проекте OpenPose [11].

Но в результате обоих подходов к решению задачи на выход модели получаем массив данных $[N \times K \times 3]$ в двумерном пространстве, где третьей размерностью идет предсказание видимости точки на изображении.

СТОИТ ВНИМАТЕЛЬНО ПЕРЕЧИТАТЬ ПОСЛЕДНИЙ АБЗАЦ.
ВОЗМОЖНО ТРЕБУЕТ ПЕРЕДЕЛКИ

4 Domain Adaptation

В современном мире машинного обучения и искусственного интеллекта способность моделей адаптироваться к новым условиям и данным стала одной из ключевых задач. Традиционные методы обучения моделей предполагают, что данные, используемые для обучения и тестирования, имеют схожие характеристики и распределения. Однако в реальных приложениях часто возникает необходимость применять модели на данных, которые существенно отличаются от тех, на которых они были изначально обучены. Это приводит к снижению точности и эффективности моделей, что ставит под угрозу их практическое применение.

Для решения данной проблемы была разработана задача доменной адаптации, целью которой является создание методов для акклиматизации модели к целевым данным, отличающимся от исходных. В рамках этой задачи были разработаны подходы и техники, направленные на уменьшение расхождений между исходным и целевым доменами, что позволяет сохранять точность и производительность моделей в новых условиях. Эти методы способствуют переносу знаний, накопленных в исходном домене, на целевой, что относит данную задачу к области методов transfer learning.

4.1 Перенос знаний

Перенос обучения (англ. transfer learning) позволил уже существующим решениям выйти за пределы первоначально заданных задач. Этот подход позволил использовать различные архитектуры для решения новых, разнообразных проблем. К примеру, это позволило перенести опыт использования трансформеров из задач обработки естественного языка в задачи компьютерного зрения.

Помимо переноса знаний между различными областями нейронных

сетей, transfer learning предоставляет возможность использовать опыт, накопленный в процессе обучения модели, для работы с новыми данными. Эти данные могут представлять собой не только новые классы в задачах классификации или кластеризации, но и иметь значительные структурные различия, такие как язык и жанр для текстовых данных или стиль и качество для изображений. Такие методы, называемые адаптацией к новым доменам данных, позволяют существенно сократить время и ресурсы, направленные на решение новых, узкоспециализированных задач.

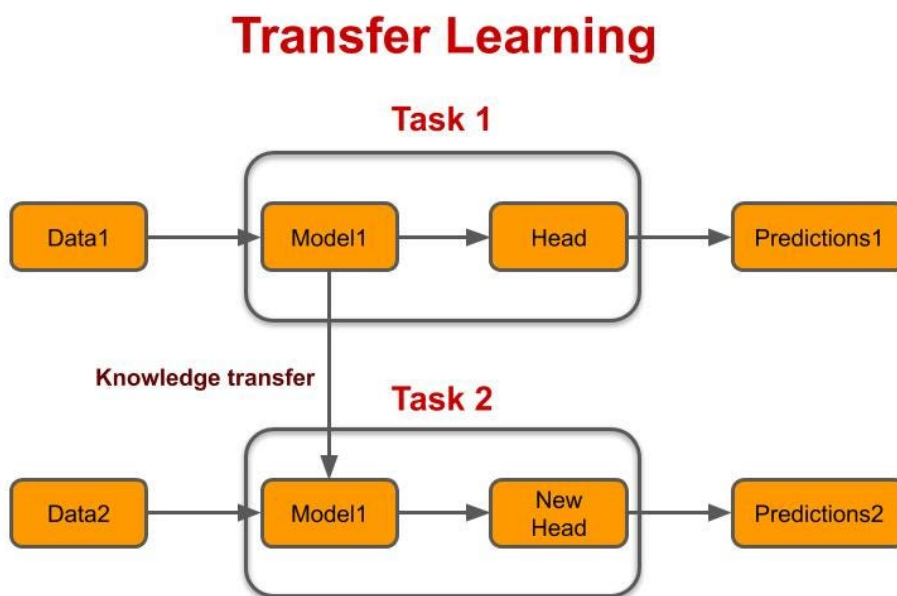


Рис. 4: Основная идея переноса знаний

ИСТОЧНИК

Основная идея состоит в том, чтобы взять модель, которая уже научилась извлекать и интерпретировать общие признаки из большого набора данных, может эффективно адаптироваться к новым данным, требующим более специфических знаний (см рис. 4). Это аналогично тому, как человек, обладая общими знаниями в одной области, может быстрее освоить смежную сферу деятельности. Таким образом получается не

только ускорить процесс обучения, но и повысить точность и эффективность модели при условии ограниченности в ресурсах.

4.2 Основные определения

Для дальнейшего повествования введем формальные обозначения для данной задачи.

Определение домена

Пусть существует пространство признаков χ и в нем задан набор данных X и частотное распределение вероятностей на нем $P(X)$:

$$X = X_1, \dots, X_n \in \chi.$$

В таком случае *доменом* D называют совокупность пространства признаков и частотного распределения вероятности на заданном наборе данных:

$$D = \{\chi, P(X)\} \quad (1)$$

Определение задачи

Пусть на пространстве признаков χ задан набор данных

$$X = X_1, \dots, X_n \in \chi. \quad (2)$$

Пусть задано пространство меток γ и для заданного набора данных существует соответственный набор меток

$$Y = \{Y_1, \dots, Y_n\} \in \gamma \quad (3)$$

Тогда *задача* T определяется как

$$T = \{Y, f(X)\}, \quad (4)$$

где f - прогностическая функция зависимости целевой переменной, которую можно рассматривать как условное вероятностное распределение $P(Y|X)$.

Определение доменной адаптации

В задаче доменной адаптации подразумевается наличие не менее двух наборов данных:

1. *Исходный домен (Source)* Представляет собой универсальный набор данных большого объема. Обозначается следующим образом:

$$D^s = \{\chi^s, P(X^s)\}$$

На домене определена задача:

$$T^s = \{Y^s, f^s(X^s)\}$$

2. *Целевой домен (Target)* Представляет собой набор данных маленького объема, к которому планируется адаптировать модель. Обозначается следующим образом:

$$D^t = \{\chi^t, P(X^t)\}$$

На домене определена задача:

$$T^t = \{Y^t, f^t(X^t)\}$$

В зависимости от того, как соотносятся между собой домены D^s и D^t и задачи T^s и T^t , ставятся различные задачи:

1. Самым распространенным является случай совпадения задач и доменов, когда $D^s = D^t$ и $T^s = T^t$. В таком случае говорят о постановке задачи классического машинного обучения. Домен D^s называют обучающей выборкой, которая используется для обучения решения

T , а домен D^t - тестовой выборкой, на которой производится оценка правильности полученного решения.

2. При совпадении данных $D^s = D^t$, но различной постановке задачи $T^s \neq T^t$ говорят о мультизадачном обучении. Часто такая постановка задачи подходит для адаптации модели классификации к новым меткам класса, которые представлены в целевой выборке, но отсутствуют в исходной.
3. При совпадении задач $T^s = T^t$, но различии в данных $D^s \neq D^t$ говорят о методах кросс-доменной адаптации. В таком случае ставится задача улучшения предсказания T^t с использованием информации, полученной в T^s .
4. При полном несовпадении данных $D^s \neq D^t$ и задач $T^s \neq T^t$ подразумевается, что пары (D, T) решают разные проблемы и адаптация происходит в индивидуальном порядке. Например, использование архитектуры трансформер, зарекомендовавшей себя в задачах обработки естественного языка, для задач компьютерного зрения.

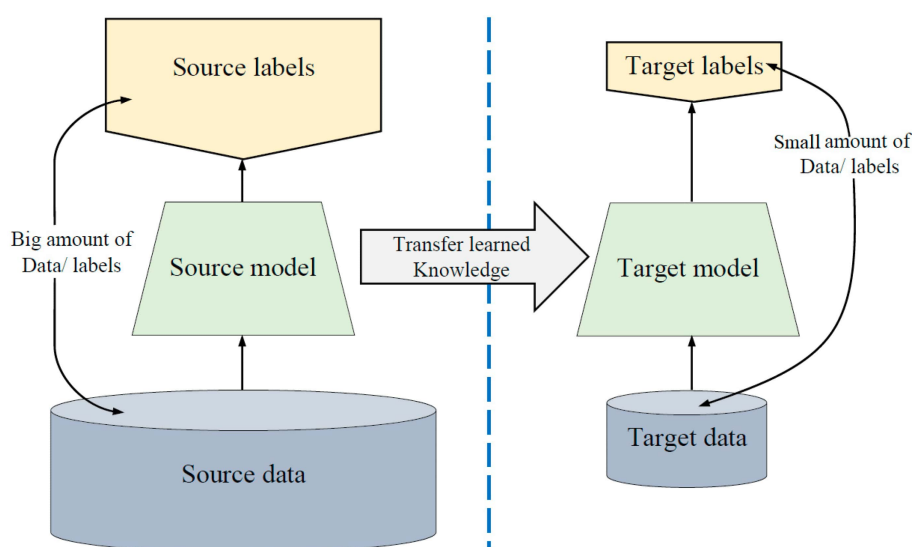


Рис. 5: Схематическое представление работы алгоритмов доменной адаптации

1. Гомогенные методы (*Homogeneous DA*)

Данный тип методов применяется, когда исходный и целевой домены имеют одинаковое пространство признаков $\chi^s = \chi^t = \chi$, но их распределения отличаются $P(X^s) \neq P(X^t)$. Методы гомогенной адаптации фокусируются на выравнивании распределений признаков между доменами, чтобы модель, обученная на исходных данных D^s , могла эффективно работать на целевых данных D^t .

2. Гетерогенные методы (*Heterogeneous DA*)

В отличие от предыдущих способов, в текущем случае отличаются и признаковые пространства между доменами $\chi^s \neq \chi^t$. Это представляет собой более сложный сценарий, так как необходимо либо преобразовывать целевые данные в такое представление, которое будет сопоставимо с исходными, либо выуживать обобщенные признаки из обоих доменов, на которых модель будет считать данные похожими.

Также стоит обратить внимание насколько сильно отличаются распределения данных, так как от этого зависит сколько шагов необходимо будет производить между доменами:

1. Одношаговая доменная адаптация (*One-step DA*)

При небольших различиях между распределениями можно использовать прямой перенос знаний из D^s в D^t . В таком случае опыт, накопленный моделью в исходном домене, напрямую используется при адаптации модели T на целевом домене. Преимущества этого метода заключаются в его простоте и скорости внедрения, так как он не требует промежуточных шагов.

2. Многошаговая доменная адаптация (*Multi-step DA*)

Однако эффективность One-Step DA может снижаться, когда различия между исходным и целевым доменами слишком велики для прямого переноса знаний. Для этих случаев и используются методы многошаговой адаптации. Они подразумевают поэтапное выполнение: знания из исходного домена сначала переносятся в один или несколько промежуточных доменов, прежде чем достигнуть целевого. Промежуточные шаги помогают постепенному выравниванию распределения данных, что улучшает адаптацию и повышает точность модели в целевом домене.

Последним маркером в классификации методов доменной адаптации является доступность размеченных данных для D^t . По данному признаку методы разделяются сразу на 4 группы:

1. *Supervised domain adaptation*

В этом случае обучение модели происходит с использованием как исходных, так и целевых данных, имеющих метки. Поэтому данные методы обычно достигают высокой точности, поскольку модель может явно учиться на целевых данных с метками.

2. *Semi-supervised domain adaptation*

К данным методам прибегают, когда в целевом домене присутствуют как размеченные данные, которые используются для начального обучения, так и неразмеченные данные, на которых есть возможность улучшать работу модели.

3. *Weakly supervised domain adaptation*

Когда данные целевого домена слабо размечены, то есть использовалась автоматическая система разметки, которая допускает ошибки, используются методы weakly supervised DA. Они направлены

либо на улучшение точности слабых меток, либо используют «мягкие» техники регуляризации, что позволяет учитывать возможность ошибки в данных.

4. *Unsupervised domain adaptation*

Самый сложный случай, когда нет возможности аннотировать домен и надо на сырых данных улучшить качество модели. Для методов этого типа часто используется кластеризация данных и итеративные подходы, которые помогают модели T уучшать собственные предсказания.

5 Обзор существующих решений

В данной главе необходимо провести анализ существующих моделей. Аргументировать выбор тех или иных моделей, методов. Подготовить теоретическую базу перед экспериментов.

5.1 Обзор моделей для распознавания КТ

DeepPose

Чисто для исторической справки. Если нужен будет объем.

OpenPose

??? год. Проект от лаборатории ??? от института ???.

Показала хорошие результаты при предыдущих сравнениях. Использует

BlazePose

Проект MediaPipe от гугл использует сеть архитектуры BazePose.

Показывает хорошие результаты. Имеет хорошие

HRNet

2019 год. Основная модель от проекта MMPose. Использует интересную архитектуру и хорошо показала себя при предыдущих сравнениях.

ViTPose

Трансформер. ??? год. Интересная архитектура. Рассказать про трансформер, рассказать про кодеки/докедеки. Рассказать про способ работы. Можно расписать неплохо.

RTMPose

2023 год. На момент исследования наиболее актуальная модель от проекта MMPose.

Работает неплохо. Обучается тоже неплохо. Попробуем заюзать в экспериментах.

Simcc + Resnet

Вообще хз что это, но если смогу, то сделаю эксперимент)

YoloPose

Несколько различных версий архитектуры Yolo дают несколько различных версий модели для распознавания позы. Использует подход снизу-вверх и хорошо

5.2 Анализ и выбор методов доменной адаптации

Опираясь на модели из предыдущей главы надо сделать анализ и выбрать только несколько методов, которые применимы к нашим требованиям и которые мы будем сравнивать между собой.

PUL

Будем это использовать в работе. Надо будет хорошо расписать...

RegDA

Интересный алгоритм. На нем базируются все остальные

НАДО БУДЕТ ОСТАВИТЬ ТОЛЬКО ОДИН ИЗ СЛЕДУЮЩИХ
МЕТОДОВ.

UDA PoseEstimation

Адаптация от синтетических данных к реальным. Должно бтыть интересно описать. Не использовали, так как до сих пор мы не перешли в 3х мерное пространство.

POST

Похожее на предыдущее

SFDA

Похожее на предыдущее

6 Эксперимент

В данной главе будет описан эксперимент, как таковой.

6.1 Описание эксперимента

Постановка эксперимента. Что планируется сделать и какие результаты хочется получить. Какие метрики будем использовать и по каким метрикам будем сравнивать.

Выбор модели для эксперимента

В рамках поставленного эксперимента поставлена задачи исследования работы алгоритма PUL для задачи распознавания ключевых точек для нескольких моделей:

1. HRNet
2. ViTPose
3. RTMPose ???
4. SimCC + ResNet ???

Предложенные модели будут обучены на исходном датасете в течение 20 эпох. Таким образом будут получены базовые наборы весов, от которых и будет проводиться дальнейшее исследование.

Описание метода доменной адаптации

Отбор точек будет производиться путем сравнения ЗНАЧЕНИЯ (ЗАМЕНИТЬ) с заранее заданным пороговым значением. ЗНАЧЕНИЕ будет отбираться двумя способами:

1. Средняя уверенность в предсказанных значениях

Для каждого предсказанного результата модель возвращает значение уверенности в своем предсказании. Усредняя это значения по всем ключевым точкам получаем среднюю уверенность для фотографии

2. Средняя уверенность для всех видимых точек

Так как уверенность на точках, которые не видно на фотографии может сильно занижать среднее ЗНАЧЕНИЕ для всего результата, то принято решение отбрасывать эти значения при вычитывании ЗНАЧЕНИЯ

По порогу уверенности будет отбираться набор значений с псевдо-разметкой, на которой модель будет дообучаться. Таким образом будет проведено N (УКАЗАТЬ ТОЧНОЕ ЗНАЧЕНИЕ) итераций адаптации и сравнено значение результатов модели при различных способах отбора псевдоразметки. Также в рамках эксперимента произведено полное дообучение модели на размеченном целевом домене и результаты будут предоставлен для сравнения с адаптированными.

Метрики оценки качества распознавания

ТАКЖЕ НЕОБХОИМО СКАЗАТЬ ОБ МЕТРИКАХ ОЦЕНКИ РАСПОЗНАВАНИЯ

Используемые ресурсы

ТАКЖЕ НЕОБХОДИМО СКАЗАТЬ ОБ ИСПОЛЬЗУЕМЫХ РЕСУРСАХ

6.2 Данные

Для эксперимента требовалось найти два различных набора данных. В качестве исходного набора был отобран датасет СОСО, который является основным для большинства современных моделей. Целевой набор был собран и размечен специально для данной задачи.

Исходные домен

Здесь надо рассказать про датасет коко и привести немного сведений о нем.

Пример ФОТОГрафий из датасета.

Рассказать про аннотацию тут.

Рассказать про отбор фотографий из-за граниченности ресурсов.

Целевой домен

Описание собранных данных. Численный объем датасета. Возможно количество локаций и распределение по ним.

Необходимо предоставить данные по распределению данных между людьми, по количеству данных для обучения/тестирования. Предоставить данные по локациям. Предоставить картинки с примерами данных, которые были собраны.

Описание системы полуавтоматической разметки данных. Что там использовалось и как проходит разметка.

Для разметки собранных данных была создана система полуавтоматической разметки данных pose-markup (ССЫЛКА). Система состоит из двух частей: автоматическая разметка ключевых точек с помощью нейросети от MediaPipe и корректировка полученных данных экспертом.

На данный момент проект ориентирован на разметку позы только на видео файлах, так как это было необходимо реализовать в рамках

эксперимента. В будущем планируется добавить возможность размечать точки и на одиночных изображениях.

Добавить коллаж примера работы системы (3 фото: исходное, после автоматической разметки, после корректировки. Необходимо явно указать точки, которые были изменены.)

Численные данные о разметке данных. Сколько пришлось разметить, сколько пришлось скорректировать. Предоставить некоторые картинки по распределению корректировки данных.

В рамках разметки данных можель имела неточности, поэтому роль эксперта была существенна при валидации работы автоматической части системы. Процент фотографий на которых были проведены изменения: собрать статистику и предоставить результаты корректировки разметки: сколько суммарно точек было изменено, сколько точек имели изменения более 5% от роста человека (метрика в бакалаврском дипломе есть), тоже самое для 15%, три случая для среднего количества точек на фотографии, три случая для фотографий (может большинство изменений были на трети фотографий, а остальные не меняли), представить сводную таблицу изменений для точек по топографии КОКО, визуализировать сводную таблицу на картинке.

6.3 Результаты эксперимента

Предоставить относительно сухо результаты эксперимента. Можно дать базовый анализ ситуации и того, что мы видим.

Необходимо предоставить результаты по времени обучения нейросетей, времени дообучения нейросетей.

Собрать данные по количеству ошибок до-после обучения. Собрать данные по количеству ошибок при изменении домена.

Данные по ресурсам, на которых обучались нейросетки.

7 Заключение

Необходимо рассказать о результатах исследования и о том, что мы получили. Привести сравнения, если таковые будут иметь место. Закинуть удочку для будущих исследований.

Список литературы

- [1] *FIFA*. Semi-automated offside technology to be used at FIFA World Cup 2022. — <https://www.fifa.com/fifaplus/en/articles/semi-automated-offside-technology-to-be-used-at-fifa-world-cup-2022>. — 2022.
- [2] Player pose analysis in tennis video based on pose estimation / Ryunosuke Kurose, Masaki Hayashi, Takeo Ishii, Yoshimitsu Aoki // 2018 International Workshop on Advanced Image Technology (IWAIT). — 2018. — Pp. 1–4.
- [3] *Thorpe, James*. Pose estimation: utilising AI to improve tennis technique. — <https://sportretina.com/blog/pose-estimation-utilising-ai-to-improve-tennis-technique/>. — 2023.
- [4] *google.github.io*. MediaPipe.Home. — <https://google.github.io/mediapipe/>.
- [5] *Kreiss, Sven*. OpenPifPaf: Composite Fields for Semantic Keypoint Detection and Spatio-Temporal Association / Sven Kreiss, Lorenzo Bertoni, Alexandre Alahi // *IEEE Transactions on Intelligent Transportation Systems*. — 2022. — Vol. 23, no. 8. — Pp. 13498–13511.
- [6] *Lin, Tsung-Yi*. Microsoft COCO: Common Objects in Context. — 2014. <https://arxiv.org/abs/1405.0312>.
- [7] *Tsung-Yi Lin Matteo Ruggero Ronchi, Alexander Kirillov*. COCO 2020 Keypoint Detection Task. — <https://cocodataset.org/#keypoints-2020>.

- [8] Human Pose Estimation Using Body Parts Dependent Joint Regressors / Matthias Dantone, Juergen Gall, Christian Leistner, Luc Van Gool // 2013 IEEE Conference on Computer Vision and Pattern Recognition. — 2013. — Pp. 3041–3048.
- [9] *Toshev, Alexander*. DeepPose: Human Pose Estimation via Deep Neural Networks / Alexander Toshev, Christian Szegedy // 2014 IEEE Conference on Computer Vision and Pattern Recognition. — 2014. — Pp. 1653–1660.
- [10] RMPE: Regional Multi-person Pose Estimation / Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, Cewu Lu // ICCV. — 2017.
- [11] OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields / Z. Cao, G. Hidalgo Martinez, T. Simon et al. // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2019.
- [12] *Contributors, MMPose*. OpenMMLab Pose Estimation Toolbox and Benchmark. — <https://github.com/open-mmlab/mmpose>. — 2020.
- [13] Detailed, Accurate, Human Shape Estimation From Clothed 3D Scan Sequences / Chao Zhang, Sergi Pujades, Michael J. Black, Gerard Pons-Moll // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — July.
- [14] Learning from Synthetic Humans / Gül Varol, Javier Romero, Xavier Martin et al. // CVPR. — 2017.
- [15] Deep 3D human pose estimation: A review / Jinbao Wang, Shujie Tan, Xiantong Zhen et al. // *Computer Vision and Image Understanding*. — 2021. — Vol. 210. — P. 103225. <https://www.sciencedirect.com/science/article/pii/S1077314221000692>.

- [16] *Tome, Denis*. Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image / Denis Tome, Chris Russell, Lourdes Agapito // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — Pp. 5689–5698.
- [17] *Whittle, Michael W.* Clinical gait analysis: A review / Michael W. Whittle // *Human Movement Science*. — 1996. — Vol. 15, no. 3. — Pp. 369–387. <https://www.sciencedirect.com/science/article/pii/0167945796000061>.
- [18] *google.github.io*. Pose Classification. — https://google.github.io/mediapipe/solutions/pose_classification.html.
- [19] OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields / Z. Cao, G. Hidalgo Martinez, T. Simon et al. // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2019.
- [20] *Kocabas, Muhammed*. MultiPoseNet: Fast Multi-Person Pose Estimation using Pose Residual Network. — 2018. <https://arxiv.org/abs/1807.04067>.
- [21] Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments / Catalin Ionescu, Dragos Papava, Vlad Olaru, Cristian Sminchisescu // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2014. — jul. — Vol. 36, no. 7. — Pp. 1325–1339.
- [22] Prediction of Human Activities Based on a New Structure of Skeleton Features and Deep Learning Model / Neziha Jaouedi, Francisco J. Perales, José Maria Buades et al. // *Sensors*. — 2020. — Vol. 20, no. 17. <https://www.mdpi.com/1424-8220/20/17/4944>.

- [23] PaStaNet: Toward Human Activity Knowledge Engine / Yong-Lu Li, Liang Xu, Xinpeng Liu et al. // CVPR. — 2020.
- [24] *Bazarevsky, Valentin*. BlazePose: On-device Real-time Body Pose tracking. — 2020. <https://arxiv.org/abs/2006.10204>.
- [25] *Bazarevsky, Valentin*. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs. — 2019. <https://arxiv.org/abs/1907.05047>.
- [26] *Zhang, Fan*. MediaPipe Hands: On-device Real-time Hand Tracking. — 2020. <https://arxiv.org/abs/2006.10214>.
- [27] *Ivan Grishchenko, Valentin Bazarevsky*. MediaPipe Holistic — Simultaneous Face, Hand and Pose Prediction, on Device. — <https://ai.googleblog.com/2020/12/mediapipe-holistic-simultaneous-face.html>.
- [28] *Sandler, Mark*. MobileNetV2: Inverted Residuals and Linear Bottlenecks. — 2019.
- [29] 2D Human Pose Estimation: New Benchmark and State of the Art Analysis / Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, Bernt Schiele // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2014. — June.
- [30] *Johnson, Sam*. Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation / Sam Johnson, Mark Everingham // Proceedings of the British Machine Vision Conference. — 2010. — doi:10.5244/C.24.12.
- [31] Hand Keypoint Detection in Single Images using Multiview Bootstrapping / Tomas Simon, Hanbyul Joo, Iain Matthews, Yaser Sheikh // CVPR. — 2017.

- [32] Convolutional pose machines / Shih-En Wei, Varun Ramakrishna, Takeo Kanade, Yaser Sheikh // CVPR. — 2016.
- [33] Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis / Laith Alzubaidi, Mohammed A. Fadhel, Omran Al-Shamma et al. // *Electronics*. — 2020. — Vol. 9, no. 3. <https://www.mdpi.com/2079-9292/9/3/427>.
- [34] Domain Adaptation: Challenges, Methods, Datasets, and Applications / Peeyush Singhal, Rahee Walambe, Sheela Ramanna, Ketan Kotecha // *IEEE Access*. — 2023. — Vol. 11. — Pp. 6973–7020.