

# Паноптическая сегментация

Александр Кириллов<sup>1, 2</sup>Каймин Хэ<sup>1</sup>Росс Гиршик<sup>1</sup>Carsten Rother<sup>2</sup>Петр Доллар<sup>1</sup><sup>1</sup> Facebook AI Research (ЯРМАРКА)<sup>2</sup> HCI / IWR, Гейдельбергский университет, Германия

## Аннотация

Мы предлагаем исследование задачи, мы называем рапортс сегменте не за горами. Паноптическая сегментация объединяет обычно не-сегментации (назначить класс метки для каждого пикселя) и экземпляра сегментирования (определение и сегмент каждого экземпляра объекта). Предлагаемая задача требует создания согласованной сегментации сцены, которая была бы богатой и полной, что является важным шагом на пути к системам реального видения. В то время как ранние работы в области компьютерного зрения решают соответствующие им- возраст/место разбора заданий, они не являются в настоящее время популярна, возможно, из-за отсутствия соответствующих показателей, или связанных с признания проблемы. Чтобы решить эту проблему, мы предлагаем новую метрику рапортс quality (PQ), которая отражает производительность для всех классов (материалов и штуквин) в интерпретируемой и унифицированной форме. Используя предложенный показатель, мы проводим тщательное исследование производительности как человека, так и машины для PS на трех существующих наборах данных, раскрывая интересную информацию о задаче. Цель нашей работы - возродить интерес сообщества к более унифицированному представлению о сегментации изображений.

## 1. Введение

На заре компьютерного зрения вещам - счетным объектам, таким как люди, животные, инструменты, - уделялась доминирующая доля внимания. Подвергая сомнению мудрость этой тенденции, Адельсон [1] повысил важность изучения систем, распознающих вещества - аморфные области со сходной текстурой или материалом, такими как трава, небо, дорога. Эта дихотомия между вещами сохраняется и по сей день, отражаясь в как разделении задач визуального распознавания, так и в специальных алгоритмах, разработанных для задач с вещами.

Изучение материала чаще всего формулируется как задача, известная как семантическая сегментация, см. Рисунок 1b. Поскольку материал аморфен и не поддается подсчету, эта задача определяется как простое присвоение метки класса каждому пикселю изображения (обратите внимание, что семантическая сегментация рассматривает классы объектов как материал). Напротив, изучение объектов обычно формулируется как задача обнаружения объекта или сегментации экземпляра, где целью является обнаружение каждого объекта и определение его границ с помощью рамки или маски сегментации соответственно, см. Рисунок 1c. Несмотря на кажущуюся взаимосвязь, наборы данных, детали и показатели

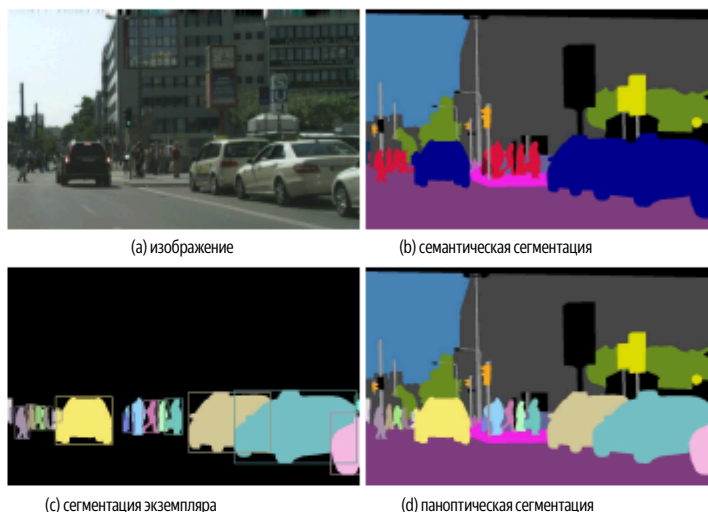


Рисунок 1: Для данного (a) изображение, мы показываем основную истину для: (b) семантической сегментации (метки классов по пикселям), (c) сегментация экземпляра (для каждого объекта маска и метка класса), и (d) предлагаемая задача паноптической сегментации (класс по пикселям + метки экземпляров). Задача PS: (1) охватывает как stuff, так и классы thing, (2) использует простой, но общий формат и (3) вводит единую оценку метрики для всех классов. Паноптическая сегментация обобщает как семантическую, так и инстанс-сегментацию, и мы ожидаем, что единая задача создаст новые проблемы и позволит использовать новые инновационные методы.

для этих двух задач визуального распознавания они существенно различаются.

Раскол между семантической сегментацией и сегментацией экземпляров привел к параллельному различию в методах решения этих задач. Stuff классификаторы обычно строятся на полностью сверточных сетях [30] с расширением [52, 5] в то время как детекторы объектов часто используют предложения object [15] и зависят от региона [37, 14]. Общий алгоритмический прогресс в решении этих задач был невероятным за последнее десятилетие, однако, что-то важное можно упустить из виду, сосредоточившись на этих задачах отдельно.

Возникает естественный вопрос: может ли быть согласование между материалом и вещами? И каков наиболее эффективный дизайн единой системы видеонаблюдения, которая генерирует богатые и последовательные сегменты сцены? Эти вопросы, в частности, особенно важны с учетом их значимости в реальном мире применений, таких как автономное вождение или дополненной реальности.

Интересно, что в то время как семантическая сегментация и сегментация экземпляров доминируют в текущей работе, в эпоху, предшествующую глубокому обучению,

был ли интерес к совместной задаче, описанной с использованием различных названий , таких как разбор сцены [42], синтаксический анализ изображений [43], или целостное понимание сцены [51]. Несмотря на свою практическую значимость, это общее направление в настоящее время не пользуется популярностью, возможно, из-за отсутствия соответствующих показателей или проблем с распознаванием.

В нашей работе мы стремимся возродить это направление. Мы предлагаем задачу, которая: (1) охватывает как материал, так и классы вещей, (2) использует простой, но общий формат вывода и (3) вводит единую метрику оценки. Чтобы четко устранить неоднозначность с предыдущей работой, мы называем результирующую задачу паноптической сегментацией (PS). Определение "паноптический" означает "включающий все видимое в одном представлении", в нашем контексте рапортis относится к единому, глобальному представлению о сегментации.

в формат задач мы применяем для рапортis сегментацию просто: каждому пикселю изображения должна быть присвоена семантическая метка и идентификатор экземпляра. Пиксели с одинаковой меткой и идентификатором принадлежат одному и тому же объекту; для меток материала идентификатор экземпляра игнорируется. Смотрите рисунок 1d для визуализации. Этот формат применялся ранее, особенно методами, которые обеспечивают неперекрывающиеся сегментации экземпляров [18, 28, 2]. Мы используем это для нашей совместной задачи, которая включает в себя разные вещи.

Фундаментальным аспектом паноптической сегментации является метрика задачи, используемая для оценки. Хотя многочисленные существующие метрики популярны либо для семантической сегментации, либо для сегментации экземпляров , эти метрики лучше всего подходят либо для материала, либо для вещей, соответственно, но не для того и другого. Мы считаем, что использование непересекающихся показателей является одной из основных причин, по которой сообщество обычно изучает сегментацию материалов изолированно. Чтобы решить эту проблему, мы вводим метрику рапортis quality (PQ) в §4. PQ прост и информативен и, что наиболее важно, может использоваться для единообразного измерения производительности как для персонала, так и для вещей. Мы надеемся, что предлагаемый совместный показатель поможет в более широком внедрении совместной задачи.

Задача паноптической сегментации включает в себя как се- мантическую, так и инстанс-сегментацию, но создает новые математические проблемы. В отличие от семантической сегментации, она требует дифференциации отдельных экземпляров объектов; это создает проблему для полностью сверточных сетей. В отличие от экземпляра сегментация, сегменты объектов не должны перекрываться ; это представляет проблему для методов, основанных на регионах, которые работают с каждым объектом независимо. Создание последовательных сегментов изображений, которые устраняют несоответствия между материалом и вещами, является важным шагом на пути к реальному использованию.

Поскольку и основная истина, и формат алгоритма для PS должны иметь одинаковую форму, мы можем выполнить подробное исследование согласованности действий человека при паноптической сегментации. Это позволяет нам более детально понять метрику PQ, включая детальную разбивку распознавания по сравнению с сегментацией и производительности "материал против вещей". Более того, измерение PQ человека помогает обосновать наше понимание эффективности машины. Это важно, поскольку позволит нам отслеживать насыщенность производительности в различных наборах данных для PS.

Наконец, мы проводим первоначальное исследование производительности машины.

руководство по PS. Для этого мы определяем простую и, вероятно, недостаточно оптимальную эвристику, которая объединяет выходные данные двух независимых систем для семантической сегментации и сегментации экземпляров посредством серии этапов постобработки, которые объединяют их выходные данные (по сути, сложная форма не максимального подавления ). Наша эвристика устанавливает базовый уровень для PS и дает нам представление об основных алгоритмических проблемах, которые она представляет.

Мы изучаем производительность как человека, так и машины на трех популярных наборах данных сегментации, которые содержат как материал , так и аннотации к вещам. Сюда входят городские пейзажи. [6], ADE20k [55], и Картографические перспективы [35] наборы данных. Для каждого из этих наборов данных мы получили результаты с использованием самых современных методов непосредственно от организаторов конкурса. В будущем мы расширим наш анализ на COCO [25] на котором материал комментируется [4]. Вместе наши результаты на этих данных сформировать прочную основу для изучения как Ху- мужчины и производительность машины на рапортis сегментации.

Оба COCO [25] и Картографические виды [35] представили задачу паноптической сегментации в качестве одного из треков в своих испытаниях по распознаванию на ECCV 2018. Мы надеемся, что включение PS в эти популярные наборы данных для распознавания наряду с треками экземпляра и семантической сегментации в этих наборах данных поможет привести к более широкому принятию предлагаемой совместной задачи.

## 2. Сопутствующая работа

Новые наборы данных и задачи играли ключевую роль на протяжении всей истории компьютерного зрения. Они помогают стимулировать прогресс и делать возможными прорывы в нашей области, и, что не менее важно, они помогают нам измерять и признавать прогресс, достигнутый нашим сообществом. Например, ImageNet [38] способствовал недавней популяризации методов глубокого обучения для визуального распознавания [20] и иллюстрирует потенциальную преобразующую силу, которой могут обладать наборы данных и задачи. Наши цели при внедрении panoptic segmentation схожи: бросить вызов нашему сообществу, стимулировать повторный поиск в новых направлениях и обеспечить как ожидаемые, так и неожиданные инновации. Далее мы рассмотрим связанные задачи. Задачи обнаружения объектов. Ранняя работа над распознаванием лиц с использованием специальных наборов данных (например, [44, 46]) помогли популяризировать обнаружение объектов с помощью ограничивающей рамки. Позже были созданы наборы данных для обнаружения пешеходов [8] помог добиться прогресса в этой области. Набор данных о PAS- CAL VOC [9] обновил задачу до более разнообразного набора общих классов объектов на более сложных изображениях. Совсем недавно набор данных COCO [25] перенесенное обнаружение на- выполняет задачу сегментации экземпляра. Сформулировав эту задачу и предоставив высококачественный набор данных, COCO помогла определить новое и захватывающее направление исследований и привела ко многим недавним прорывам в сегментации экземпляров [36, 24, 14]. Наши общие цели для паноптической сегментации схожи. Задачи семантической сегментации. Семантическая сегментация наборы данных имеют богатую историю [39, 26, 9] и помогли внедрить ключевые инновации (например, полностью сверточные сети [30] были разработаны с использованием [26, 9]). Эти наборы данных содержат и то, и другое.

и дело классов, но не различают отдельного объекта в позицию. В последнее время в области наблюдается множество новых сегментов не за горами.

- ции данных, в том числе городские пейзажи [6], ADE20k [55], и Картографические перспективы [35]. Эти наборы данных фактически поддерживают как семантическую сегментацию, так и сегментацию экземпляров, и каждый из них предпочел иметь отдельный трек для двух задач. Важно отметить, что они содержат всю информацию, необходимую для PS. Другими словами, задача panoptic segmentation может быть запущена на этих наборах данных без какого-либо нового сбора данных.

Многозадачное обучение. Благодаря успеху глубокого обучения для многих задач визуального распознавания возник значительный интерес к многозадачным подходам к обучению, которые обладают широкой полезностью и могут решать множество разнообразных проблем со зрением в единой структуре [19, 32, 34]. Например, UberNet [19] решает множество визуальных задач от низкого до высокого уровня, включая определение объектов и семантическую сегментацию, используя единую сеть. Несмотря на значительный интерес к этой области, мы подчеркиваем, что паноптическая сегментация - это не многозадачная задача, а скорее единый, унифицированный подход к сегментации изображений. Ка- Чески, в многозадачном режиме настройка позволяет независимо и по- сообщать несогласованных выходов за вещи и вещи, а PS требует единого согласованного места сегментации. Совместные задачи сегментации. В эпоху, предшествующую глубокому обучению, был значительный интерес к созданию последовательных интерпретаций сцен . Основополагающая работа по анализу изображений [43] предложил общую байесовскую структуру для совместного моделирования сегментации, обнаружения и распознавания. Позже, подходы, основанные на графических моделях, изучают согласованную сегментацию материалов и вещей [51, 41, 42, 40]. Хотя эти методы общей мотивации, нет согласованных задач Опре- определению, и различных выходных форматов и разной оценки показателей были использованы, в том числе отдельные показатели для оценки результатов на вещи, и вещь классов. В последние годы это направление стало менее популярным, возможно, по этим причинам.

В нашей работе мы стремимся возродить это общее направление, но, в отличие от предыдущих работ, мы фокусируемся на самой задаче. Конкретно, как обсуждалось, PS: (1) обращается как к материалу, так и к классам thing , (2) использует простой формат и (3) вводит метрику единой формы как для материала, так и для вещей. Предыдущая работа по совместной сегментации использовала различные форматы и непересекающиеся методы для оценки материала и сущностей. Методы, которые генерируют, которые не перекрывают друг друга, сегментации экземпляров. Методы, которые генерируют непересекающиеся сегментации экземпляров [18, 3, 28, 2] используйте тот же формат, что и PS, но эти методы обычно относятся только к классам одежды. Обращаясь как к материалу, так и к вещам, используя простой формат и вводя единую метрику, мы надеемся стимулировать более широкое внедрение совместной задачи.

Задача модальной сегментации. В [56] объекты снабжены комментариями по модальности: отмечен полный объем каждой области, а не только видимая. Наша работа сосредоточена на сегментации всех видимых областей, но расширение паноптической сегментации на амодальную настройку является интересным направлением для будущей работы.

### 3. Паноптический формат сегментации

Формат задачи. Формат паноптической сегментации определить просто. Учитывая заранее определенный набор классов, закодированных  $L := \{0, \dots, L-1\}$  семантический  $L$ , задача требует алгоритм паноптической сегментации для сопоставления каждого пикселя  $i$  изображения с парой  $(l_i, z_i) \in L \times \mathbb{N}$  где  $l_i$  представляет семантический класс пикселей  $i$  и  $z_i$  представляет его экземпляр id. Символ  $z_i$  группирует пиксели одного и того же класса в отдельные сегменты . Примечания к основной правде кодируются идентично. Неоднозначным или внеклассовым пикселям может быть присвоена специальная пустая метка; т.е. Не все пиксели должны иметь семантическую метку. Материал и метки вещей. Набор семантических меток состоит из подмножеств  $L^{st}$  и  $L^{th}$ , такой , что  $L = L^{st} \cup L^{th}$  и  $L^{st} \cap L^{th} = \emptyset$ . Эти подмножества соответствуют меткам stuff и thing соответственно. Когда пиксель помечен символом  $l_i \in L^{st}$ , соответствующий ему идентификатор экземпляра  $z_i$  не имеет значения. То есть для классов материала все пиксели принадлежат одному экземпляру (например, одному и тому же небу). В противном случае все пиксели с одинаковым  $(l_i, z_i)$  как- обозначение, где  $l_i \in L^{th}$ , принадлежат одному и тому же экземпляру (например, одному и тому же автомобилю), и наоборот, все пиксели, принадлежащие одному экземпляру, должны быть одинаковыми . Выбор  $(l_i, z_i)$  какие классы являются материалом, а какие вещами - выбор дизайнера остается за создателем набора данных, как и в предыдущих наборах данных. Связь с семантической сегментацией. Формат задачи PS является строгим обобщением формата семантической сегментации. Действительно, обе задачи требуют, чтобы каждому пикселю в изображении была присвоена семантическая метка. Если основная истина не указывает экземпляры или все классы являются материалом, то форматы задач идентичны (хотя метрики задач различаются). Кроме того, включение классов объектов, которые могут иметь множество экземпляров для каждого изображения, дифференцирует задачи. Связь с сегментацией экземпляров. Задача сегментации экземпляра требует наличия метода для сегментации каждого объекта в изображении. Однако это допускает перекрытие сегментов, тогда как задача паноптической сегментации разрешает присваивать каждому пикселю только одну семантическую метку и один идентификатор экземпляра. Следовательно, для PS при построении перекрытия невозможны. В следующем разделе мы покажем, что это различие играет важную роль в оценке производительности. Баллы достоверности. Как и при семантической сегментации, но в отличие от сегментации экземпляров, нам не требуются баллы достоверности, связанные с каждым сегментом для PS. Это делает паноптическую задачу симметричной по отношению к людям и машинам : и те, и другие должны генерировать один и тот же тип объявления изображения . Это также упрощает оценку согласованности действий человека для PS . Это контрастирует с сегментацией экземпляров, которая нелегко поддается такому исследованию, поскольку люди-аннотаторы не предоставляют явных показателей достоверности (хотя может быть измерена единственная точка точности / отзыва). Мы отмечаем, что баллы достоверности дают нижестоящим системам больше информации, которая может быть полезной, поэтому все же может быть желательно, чтобы алгоритм PS генерировал баллы достоверности при определенных настройках.

#### 4. Метрика паноптической сегментации

В этом разделе мы представляем новую метрику для паноптической сегментации. Начнем с того, что отметим, что существующие метрики специализированы либо для семантической сегментации, либо для сегментации экземпляров и не могут быть использованы для оценки совместной задачи, включающей как stuff, так и классы thing. Предыдущая работа по совместной сегментации позволила обойти эту проблему, оценив персонал и его эффективность с использованием независимых показателей (например, [31, 47, 42, 46]). Однако это создает проблемы при разработке алгоритма-

это затрудняет сравнения и затрудняет коммуникацию.

Мы надеемся, что введение единой метрики для материала и вещей будет способствовать изучению единой задачи.

Прежде чем перейти к дальнейшим деталям, мы начнем с определения следующих желаемых параметров для подходящей метрики для PS:

Полнота. Метрика должна единообразно

относиться к классам stuff и thing, охватывая все аспекты задачи.

Интерпретируемость. Мы ищем показатель с идентифицируемым значением, который облегчает общение и понимание.

Простота. Кроме того, показатель должен быть простым для определения и реализации. Это повышает прозрачность и позволяет упростить повторную реализацию. В связи с этим метрика должна быть эффективной в вычислениях, чтобы обеспечить быструю оценку.

Руководствуясь этими принципами, мы предлагаем новую метрику panoptic quality (PQ). PQ измеряет качество прогнозируемой паноптической сегментации относительно основной достоверности. Это включает в себя два этапа: (1) сопоставление сегментов и (2) вычисление PQ с учетом совпадений. Далее мы описываем каждый шаг, затем возвращаемся к сравнению с существующими показателями.

##### 4.1. Сопоставление сегментов

Мы указываем, что предсказанный сегмент и базовый истинный сегмент могут совпадать, только если их пересечение по объединению (IoU) строго больше 0.5. Это требование вместе с свойством неперекрываемости паноптической сегментации дает уникальное соответствие: может быть не более одного предварительно продиктованного сегмента, соответствующего каждому основному сегменту истинности.

Теорема 1. Учитывая паноптическую сегментацию изображения с прогнозируемой и достоверной достоверностью, каждый сегмент достоверности может иметь не более одного соответствующего прогнозируемого сегмента с IoU строго больше 0,5 и наоборот.

Доказательство. Пусть  $g$  - основной отрезок истинности, а  $p_1$  и  $p_2$  быть двумя предсказанными сегментами. По определению,  $p_1 \cap p_2 = \emptyset$  (они не перекрываются). Поскольку  $|p_i \cup g| \geq |g|$ , мы получаем следующее:

$$(p_i, g) = \text{Долговая расписка} \leq \frac{|p_i \cap g|}{|g|} \quad \text{для } i \in \{1, 2\}.$$

Суммируя над  $i$  и поскольку тот факт, что  $p_1 \cap p_2 = \emptyset$  из-за  $|p_i \cap g| + |p_j \cap g| \leq |g|$ , получаем:

$$\text{Долговая расписка}(p_1, g) + \text{Долговая расписка}(p_2, g) \leq \frac{|p_1 \cap g| + |p_2 \cap g|}{|g|} \leq 1.$$

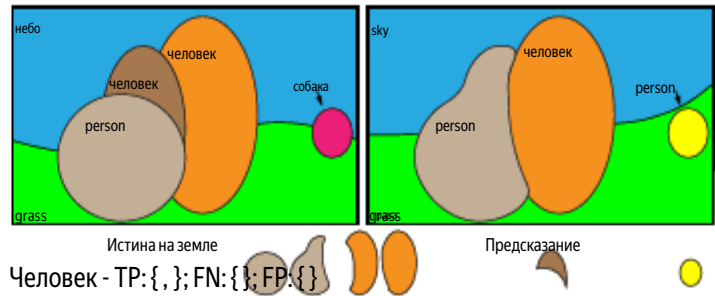


Рисунок 2: Игрушечная иллюстрация основной истины и предсказанных паноптических сегментов изображения. Пары сегментов одного цвета имеют размер расписки больше 0,5 и, следовательно, совпадают. Мы покажем, как сегменты для класса person разделяются на истинно положительные результаты TP, ложноотрицательные результаты FN и ложноположительные результаты FP.

Следовательно, если долговая расписка  $(p, g) > 0.5$ , тогда долговая расписка  $(p, g)$  должно быть меньше 0.5. Можно использовать обратное изменение роли  $p$  и  $g$ , чтобы доказать, что только один сегмент истинности может иметь долговую расписку с предсказанным сегментом, строго большим 0,5.  $\square$

Требование о том, что совпадения должны иметь значение IoU больше 0,5, что, в свою очередь, приводит к теореме об уникальном совпадении, обеспечивает два наших желаемых свойства. Во-первых, это просто и эффективно, поскольку соответствия уникальны и тривиальны для получения. Во-вторых, он поддается интерпретации и легок для понимания (и не требует решения сложной задачи сопоставления, как это обычно бывает для этих типов показателей [13, 50]).

Обратите внимание, что из-за свойства уникальности, для  $\text{IoU} > 0.5$ , любая разумная стратегия сопоставления (включая жадную и оперативную) даст идентичное сопоставление. Для меньшего количества IoU потребуются другие методы сопоставления; однако в экспериментах мы покажем, что более низкие пороговые значения не являются необходимыми, поскольку совпадения с  $\text{IoU} \leq 0.5$  на практике редки. 4.2. Вычисление PQ

Мы рассчитываем PQ для каждого класса независимо и усредняем возраст по классам. Это делает PQ нечувствительным к балансу классов. Для каждого класса уникальное сопоставление разбивает предсказанные и достоверные сегменты на три набора: истинные положительные результаты (TP), ложноположительные результаты (FP) и ложноотрицательные результаты (FN), представляющие сопоставленные пары сегментов, не имеющих аналогов (предсказанные сегменты и непревзойденные достоверные сегменты, соответственно). Пример проиллюстрирован на рисунке 2. Учитывая эти три набора, PQ определяется как:

$$PQ = \frac{\text{Долговая расписка}(p, g) \mid (p, g) \in TP}{|TP| + |FP| + |FN|} \quad (1)$$

PQ интуитивно понятен после проверки:  $\frac{1}{|TP|}$   $(p, g) \in TP$  Долговая расписка  $(p, g)$

- это просто средняя долговая расписка сопоставленных сегментов, в то время как  $|FP| + |FN|$  добавляется к знаменателю для наказания сегменты без совпадений. Обратите внимание, что все сегменты имеют одинаковую важность независимо от их площади. Более того, если мы умножим и разделим PQ на размер набора TP, то

PQ можно рассматривать как умножение количества (SQ) члена и качества распознавания (RQ) термин:

$$PQ = \frac{(p, g)_{\text{ст}}}{|TP|} \times \frac{(p, g) \text{ Долговая расписка}}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} \quad (2)$$

качество сегментации (SQ)                      качество распознавания (RQ)

Написанный таким образом, RQ является знакомым значением  $F_1$  оценка [45] широко используется для оценки качества в настройках обнаружения. [33]. SQ - это просто среднее значение IoU сопоставленных сегментов. Мы находим разложение  $PQ = SQ \times RQ$ , чтобы обеспечить понимание для анализа. Отметим, однако, что эти два значения не являются независимыми, поскольку SQ измеряется только по согласованным сегментам.

Наше определение PQ соответствует нашим желаниям. Он измеряет производительность всех классов единообразным способом, используя простую и интерпретируемую формулу. В заключение мы обсудим, как мы обрабатываем пустые области и группы экземпляров. [25].

Пустые метки. В основной правде есть два источника пустых меток: (а) пиксели вне класса и (б) неоднозначные или неизвестные пиксели. Поскольку часто мы не можем отличить эти два случая, мы не оцениваем прогнозы для пустых пикселей. В частности, : (1) во время сопоставления все пиксели в прогнозируемом сегменте, которые помечены как недействительные в основной правде, удаляются из прогнозирования и не влияют на вычисление IoU, и (2) после сопоставления, несогласованные предсказанные сегменты, которые содержат часть пустых пикселей, превышающих пороговое значение соответствия, повторно перемещаются и не считаются ложноположительными. Наконец, выходные данные также могут содержать пустые пиксели; это не влияет на оценку.

Группируйте метки. Обычная практика создания аннотаций [6, 25] заключается в использовании метки группы вместо идентификаторов экземпляров для смежных позиций одного и того же семантического класса, если точное определение каждого экземпляра затруднено. Для вычисления PQ: (1) во время сопоставления области группы не используются, и (2) после сопоставления несогласованные предсказанные сегменты, которые содержат долю пикселей из группы того же класса, превышающую пороговое значение сопоставления, удаляются и не считаются ложноположительными.

#### 4.3. Сравнение с существующими показателями

В заключение мы сравниваем PQ с существующими метриками для семантической сегментации и сегментации экземпляров. Метрики семантической сегментации. Общие показатели для семантической сегментации включают точность пикселей, точность среднего значения и количество долговых расписок. [30]. Эти показатели вычисляются только на основе выходных данных / меток пикселей и полностью игнорируют метки на уровне объекта. Например, IoU - это отношение между правильно заданными продиктованными пикселями и общим количеством пикселей либо в прогнозе, либо в основной правде для каждого класса. Поскольку эти показатели игнорируют метки экземпляров, они не очень хорошо подходят для оценки классов thing. Наконец, обратите внимание, что значение IoU для семантической сегментации отличается от нашего качества сегментации (SQ), которое вычисляется как среднее значение IoU для сопоставленных сегментов.

Показатели сегментации экземпляра. Стандартная метрика для сегментации экземпляра - средняя точность (AP). [25, 13]. AP требует, чтобы каждый сегмент объекта имел оценку достоверности

для оценки кривой точности / отзыва. Обратите внимание, что, хотя показатели достоверности вполне естественны для обнаружения объектов, они не используются для семантической сегментации. Следовательно, AP не может быть использован для измерения результатов семантической сегментации или аналогично PS (см. Также Обсуждение конфиденциальных данных в §3).

Паноптическое качество. PQ обрабатывает все классы (stuff и прочее) единообразно. Отметим, что при декомпозиции PQ

в SQ и RQ полезно интерпретировать результаты, PQ - это не комбинация показателей семантической сегментации и сегментации экземпляров. Скорее, SQ и RQ вычисляются для каждого класса (stuff и вещи) и измеряют качество сегментации и распознавания соответственно. Таким образом, PQ объединяет оценку по всем классам. Мы подтверждаем это утверждение строгой экспериментальной оценкой PQ в 7, включая сравнения с IoU и AP для семантической сегментации и сегментации экземпляра соответственно.

### 5. Наборы данных Panoptic для сегментации.

Насколько нам известно, только три общедоступных набора данных имеют как плотные семантические аннотации, так и аннотации сегментации экземпляров: Городские пейзажи [6], ADE20k [55], и Картографические перспективы [35]. Мы используем все три набора данных для паноптической сегментации. Кроме того, в будущем мы расширим наш анализ на COCO [25] на какие материалы были недавно даны комментарии [41].

Городские пейзажи [6] содержит 5000 изображений (2975 train, 500 val и 1525 test) эгоцентрических сценариев вождения в городских условиях. Он содержит плотные пиксельные аннотации (97% покрытия) 19 классов, среди которых 8 имеют сегментирование на уровне экземпляра.

ADE20k [55] содержит более 25 тыс. изображений (20 тыс. train, 2 тыс. val, 3 тыс. test), которые снабжены плотными комментариями с помощью набора меток с открытым словарем. Для конкурса Places Challenge 2017<sup>2</sup>, выбираются классы 100 thing и 50 вещей, которые занимают 89% всех пикселей. Мы используем этот закрытый словарь в нашем исследовании.

Картографические перспективы [35] содержит 25 тыс. изображений с видом на улицу (18 тыс. поезд, 2 тыс. вал, 5 тыс. тест) в широком диапазоне разрешений. "Исследовательская версия" набора данных плотно аннотирована (98% пиксельного покрытия) 28 материалами и 37 классами объектов.

### 6. Исследование согласованности действий человека

Одним из преимуществ panoptic segmentation является то, что она позволяет измерять согласованность аннотаций человека. Помимо того, что это интересно как самоцель, исследования согласованности действий человека позволяют нам понять задачу в деталях, включая детали предлагаемой нами метрики и разбивку человеческой согласованности по различным осям. Это дает нам представление о внутренних проблемах, связанных с задачей, не искажая наш анализ алгоритмическим выбором. Кроме того, исследования, проводимые человеком, помогают повысить производительность машины (обсуждается в 7) и позволяют нам откалибровать наше понимание задачи.

<sup>1</sup> Сегментации экземпляров COCO содержат перекрытия. Мы собрали глубину упорядочения для всех пар перекрывающихся экземпляров в COCO, чтобы устранить эти наложения: <http://cocodataset.org/#panoptic-2018>

<sup>2</sup> <http://placeschallenge.csail.mit.edu>





Рисунок 3: Недостатки сегментации. Изображения увеличены и обрезаны. Верхний ряд (изображение перспективы): оба аннотатора идентифицируют объект как автомобиль, однако один разделяет автомобиль на две машины. Нижний ряд (изображение городских пейзажей): сегментация действительно неоднозначна.

	PQ	PQ <sup>ST</sup>	PQ <sup>th</sup>	K <sub>th</sub>	K <sub>th</sub> <sup>ST</sup>	K <sub>th</sub> <sup>th</sup>	RQ	RQ <sup>ST</sup>	RQ <sup>th</sup>
Городские пейзажи	69.7	71.3	67.4	84.2	84.4	83.9	82.1	83.4	80.2
ADE20k	67.1	70.3	65.9	85.8	85.5	85.9	82.1	82.4	76.4
Перспективы	57.5	62.6	53.4	79.5	81.6	77.9	71.4	76.0	67.7

Таблица 1: Согласованность между людьми в отношении материала и вещей. Паноптика, сегментация и качество распознавания (PQ, SQ, RQ), усредненные по классам (PQ = SQ × RQ для каждого класса), представлены в процентах. Возможно, удивительно, но мы обнаруживаем, что согласованность действий человека в каждом наборе данных относительно одинакова как для материала, так и для вещей.

Человеческие аннотации. Для обеспечения адекватной согласованности с человеком-да, создатели dataset любезно предоставили нам 30 изображений с двойными комментариями для Cityscapes, 64 для ADE20k и 46 для Vistas. Для городских пейзажей и перспектив изображения независимо помечаются разными аннотаторами. ADE20k аннотируется одним хорошо обученным аннотантом, который помечает один и тот же набор изображений с интервалом в шесть месяцев. Чтобы измерить надежное паноптическое качество (PQ) для людей-аннотаторов, мы рассматриваем одну аннотацию к каждому изображению как основную истину, а другую как предсказание. Обратите внимание, что PQ симметричен относительно основной истины и предсказания, поэтому порядок не имеет значения.

Согласованность с человеком. Во-первых, таблица 1 показывает соответствие персонала- для каждого набора данных наряду с разложением PQ на качество сегментации (SQ) и качество распознавания (RQ). Как и ожидалось, люди не совершенны в выполнении этой задачи, что согласуется с исследованиями качества аннотаций, проведенными [6, 55, 35]. Визуализация ошибок сегментации и классификации людей показана на рисунках. 3 и 4, соответственно.

Отметим, что таблица 1 устанавливает меру согласия аннотатора с каждым набором данных, а не верхнюю границу согласованности для человека. Мы также подчеркиваем, что цифры не сопоставимы между наборами данных и не должны использоваться для оценки качества набора данных. Количество классов, процент аннотированных пикселей и сложность сцены различаются в разных наборах данных, каждый из которых существенно влияет на сложность аннотирования.

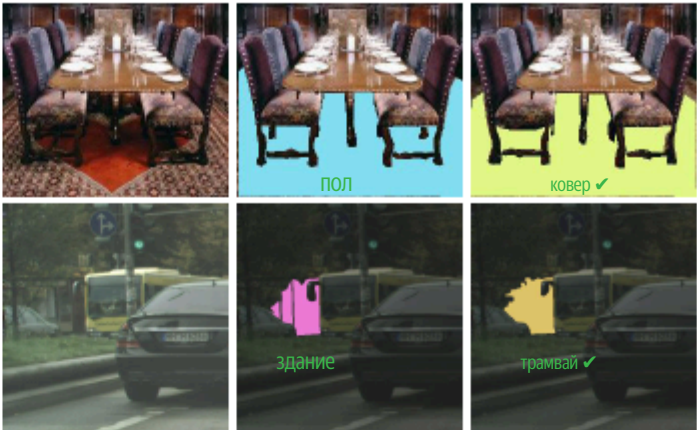


Рисунок 4: Недостатки классификации. Изображения увеличены и обрезаны. Верхняя строка (изображение ADE20k): простая неправильная классификация. Нижний ряд (изображение городского пейзажа): сцена чрезвычайно сложная, трамвай - это правильный класс для сегмента. Многие ошибки трудно устранить.

	PQ <sub>s</sub>	PQ <sub>m</sub>	PQ <sub>l</sub>	K <sub>s</sub>	K <sub>m</sub>	K <sub>l</sub>	RQ <sub>s</sub>	RQ <sub>m</sub>	RQ <sub>l</sub>
Городские пейзажи	35.1	62.3	84.8	67.8	81.0	89.9	51.5	76.5	94.1
ADE20k	49.9	69.4	79.0	78.0	84.0	87.8	64.2	82.5	89.8
Перспективы	35.6	47.7	69.4	70.1	76.6	83.1	51.5	62.3	82.6

Таблица 2: Согласованность действий человека в зависимости от масштаба для малых (S), средних (M) и крупных (L) объектов. Масштаб играет большую роль в определении согласованности действий человека при паноптической сегментации. Для крупных объектов и SQ, и RQ превышают 80 во всех наборах данных, в то время как для небольших объектов RQ резко падает. SQ для небольших объектов вполне разумно.

Материал . вещи. vs PS требует сегментации обоих материалов и прочее. В таблице 1 мы также показываем PQ<sup>ST</sup> и PQ<sup>th</sup> который является ли PQ усредненным по классам материалов и классам вещей, соответственно. Что касается городских пейзажей и человеческой согласованности ADE20k, то в отношении материала и вещей они близки, в отношении перспектив разрыв немного больше. В целом, это подразумевает, что материал и вещи имеют схожую сложность, хотя классы вещей несколько сложнее. На рисунке 5 мы показываем PQ для каждого класса в каждом наборе данных, отсортированном по PQ. Обратите внимание, что материалы и классы things распределены довольно равномерно. Это означает, что предлагаемая метрика обеспечивает хороший баланс и, действительно, успешно объединяет материал и задачи сегментации без доминирования ошибки. Маленькие и большие объекты. Чтобы проанализировать, как изменяется PQ в зависимости от размера объекта, мы разделяем наборы данных на малые (S), средние (M) и большие (L) объекты, рассматривая наименьшие 25%, средние 50% и наибольшие 25% объектов в каждом наборе данных, соответственно. В таблице 2, мы видим, что для больших объектов согласованность между людьми для всех наборов данных довольно хорошая. Для небольших объектов RQ значительно снижается, что означает, что людям-аннотаторам часто бывает трудно находить небольшие объекты. Однако, если обнаружен небольшой объект, он сегментируется относительно хорошо. Пороговое значение для IoU. Применяя перекрытие, превышающее 0,5 IoU, мы получаем уникальное соответствие по Теореме 1. Как вообще, является ли порог 0,5 разумным? Альтернативная стратегия заключается в том, чтобы не использовать пороговое значение и выполнить сопоставление путем решения задачи о двудольном сопоставлении с максимальным весом [47]. The

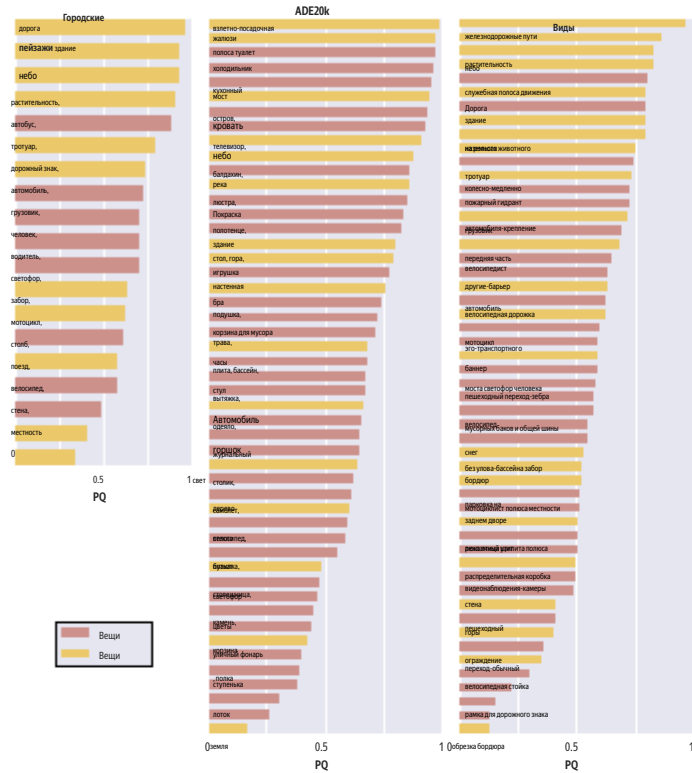


Рисунок 5: согласованность между людьми для каждого класса, отсортированная по PQ. Классы Thing выделены красным, классы stuff - оранжевым (для ADE20k показаны все остальные классы, классы, не имеющие совпадений в наборах тестов с двойными аннотациями, опущены). Вещи распределены довольно равномерно, что подразумевает, что PQ уравнивает их производительность.

оптимизация вернет соответствие, которое максимизирует сумму долговых расписок сопоставленных сегментов. Мы выполняем сопоставление, используя эту оптимизацию, и строим график кумулятивной плотности функций совпадающих перекрытий на рисунке 6. Менее чем в 16% совпадений долговые расписки пересекаются менее чем на 0,5, указывает на то, что ослабление порога должно иметь незначительный эффект.

Чтобы проверить эту интуицию, на рисунке 7 мы показываем PQ, вычисленный для различных пороговых значений долговых расписок. ..... Примечательно, что разница в PQ для IoU 0,25 и 0,5 относительно невелика, особенно по сравнению с разрывом между IoU 0,5 и 0,75, где изменение PQ больше. Более того, многие совпадения по более низким долговым распискам являются ложными совпадениями. Следовательно, учитывая, что соответствие для IoU номиналом 0,5 не только уникально, но также просто и интуитивно, мы считаем, что значение 0,5 по умолчанию является разумным. Соотношение SQ и RQ. Наше определение RQ эквивалентно  $F_1$  оценке. Однако возможны и другие варианты. Вдохновленный обобщенным  $F_\beta$  оценкой [45], мы можем ввести параметр  $\beta$ , который позволяет настраивать штраф за ошибки распознавания:

$$RQ = \frac{|TP|}{|TP| + \alpha|FP| + \alpha|FN|} \quad (3)$$

По умолчанию  $\alpha$  равно 0,5. Уменьшение  $\alpha$  уменьшает штраф за несовпадающие сегменты и, таким образом, увеличивает RQ (на SQ это не влияет). Поскольку  $PQ = SQ \times RQ$ , это изменяет относительный эффект

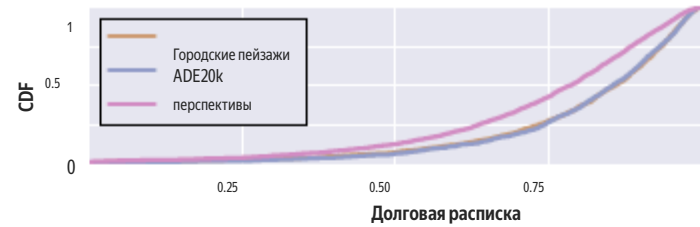
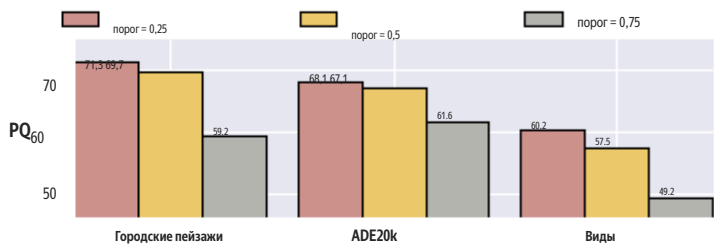


Рис. 6: Кумулятивной функции плотности перекрытий за соответствием сегментов в трех наборах данных, когда матчи вычисляются по солв-Инг максимальный взвешенный двудольный проблемы соответствия [47]. После сопоставления менее чем у 16% сопоставленных объектов количество долговых расписок ниже 0,5.



Согласованность с человеком для разных пороговых значений IoU. Рисунок 7: Разница в PQ при использовании соответствующего порога 0,25 против 0,5 относительно невелика. Для IoU, равного 0,25, соответствие получается путем решения задачи о двудольном сопоставлении с максимальным весом. Для порогового значения, превышающего 0,5, соответствие уникально, и его намного проще получить.

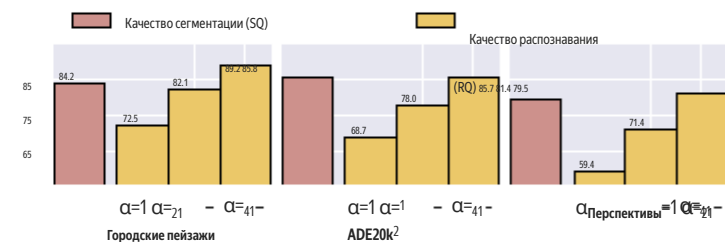


Рисунок 8: Площадь против RQ для разных  $\alpha$ , см. (3). Уменьшение  $\alpha$  уменьшает штраф за несовпадающие сегменты и, таким образом, увеличивает заявленный RQ (на SQ это не влияет). Мы повсюду используем  $\alpha$ , равное 0,5, но, настроив  $\alpha$ , можно сбалансировать влияние SQ и RQ в конечной метрике.

зависимости PS от RQ для конечной метрики PQ. На рисунке 8 мы показываем SQ и RQ для различных  $\alpha$ . Значение  $\alpha$  по умолчанию обеспечивает хороший баланс между SQ и RQ. В принципе, изменение  $\alpha$  может быть использовано для уравнивания влияния ошибок сегментации и распознавания на конечную метрику. В аналогичном духе можно также добавить параметр  $\beta$ , чтобы сбалансировать влияние FPS и FNs.

## 7. Базовые показатели производительности машины.

Теперь мы представляем простые машинные базовые линии для рапортс сегментации. Нас интересуют три вопроса: (1) Как эвристические комбинации высокопроизводительных инстансов и систем семантической сегментации работают при паноптической сегментации? (2) Как PQ соотносится с существующими показателями, такими как AP и IoU? (3) Как машинные результаты соотносятся с результатами, полученными человеком, которые мы представили ранее?

Городские пейзажи	AP		PQ <sub>н</sub>	Кв. <sub>н</sub>	RQ <sub>н</sub>	
Маска R-CNN+ COCO [14]	36.4	31.5	33.1	54,0	79,4	67.8
Маска R-CNN [14]			28.0	49,6	78,7	63.0
ADE20k	АП	АП <sub>нет</sub>	Кв.М <sub>н</sub>	Кв. <sub>нс.</sub>	RQ <sub>н</sub>	
Megvii [31]	30.1	24.8	41.1	81.6	49.6	
G-RMI [10]	24.6	20.6	35.3	79.3	43.2	

Таблица 3: Машинные результаты сегментации экземпляров (классы материала игнорируются). Непересекающиеся прогнозы получены с использованием предложенной эвристики. AP<sup>ст</sup> является AP из неперекрывающихся прогнозов. Как и ожидалось, устранение перекрытий вредит AP, поскольку детекторы выигрывают от прогнозирования нескольких перекрывающихся гипотез. Методы с лучшей точкой доступа также имеют лучшую точку доступа<sup>н</sup> и аналогично улучшен PQ.

Алгоритмы и данные.

Мы хотим понять panoptic.

сегментация с точки зрения существующих хорошо зарекомендовавших себя методов. Поэтому мы создаем базовую систему PS, применяя разумные эвристики (описанные вкратце) к выходным данным существующих систем верхнего уровня и семантической сегментации.

Мы получили выходные данные алгоритма для трех наборов данных. Для городских пейзажей мы используем выходные данные набора значений, сгенерированные ведущими в настоящее время алгоритмами (PSPNet [54] и Маска R-CNN [14] для семантической сегментации и сегментации экземпляров соответственно). Для ADE20k мы получили результаты для победителей обоих семантических [12, 11] и экземпляр [31, 10] сегментация отслеживает подмножество тестовых изображений объемом 1 тыс. из конкурса Places 2017 года. Для программы Vistas, которая используется для LSUN'17 Segmentation Challenge, организаторы предоставляют нам 1 тыс. тестовых изображений и результаты работ-победителей для треков instance и semantic segmentation [29, 53].

Используя эти данные, мы начинаем с анализа PQ для задач экземпляра и семантической сегментации по отдельности, а затем рассматриваем задачу полной паноптической сегментации. Обратите внимание, что наши "базовые линии" очень эффективны и что более простые базовые линии могут быть более разумными для объективного сравнения в статьях по PS. Сегментация экземпляров. Алгоритмы сегментации экземпляров создают перекрывающиеся сегменты. Чтобы измерить PQ, мы должны сначала устранить эти перекрытия. Для этого мы разрабатываем сим- дям не максимального подавления (НМС)-вроде порядок. Сначала мы отсортируем прогнозируемые сегменты по их показателям достоверности и удалим экземпляры с низкими показателями. Затем мы перебираем отсортированные экземпляры, начиная с наиболее достоверных. На каждый экземпляр мы удалим пикселей, которые были назначены в предыдущие сегменты, а затем, если достаточная часть Сегмента по-прежнему, мы принимаем неперекрывающиеся части, другим образом отбросить весь сегмент. Все пороговые значения выбраны поиском по сетке для оптимизации PQ. Результаты для Cityscapes и ADE20k приведены в таблице 3 (Vistas опущен, поскольку в нем была только одна запись для вызова экземпляра 2017 года). Самое важное, что AP и PQ отслеживаются тесно, и мы ожидаем, что улучшения в AP детектора также улучшат его PQ.

Семантическая сегментация. Семантические сегментации не имеют перекрывающихся сегментов по дизайну, и поэтому мы можем

Городские пейзажи	Долговая	RO <sup>ст</sup>	Кв <sup>ст</sup>	RO <sup>ст</sup>
Многомасштабная сеть PSPNet [54]	расписка 80.6	66.6	82.2	79.3
Одномасштабная сеть PSPNet [54]	79.6 Долговая	65.2	81.6	78.0
ADE20k	расписка	ПК <sup>ст</sup>	Кв.М <sup>ст</sup>	RQ <sup>ст</sup>
КАСИЯ ИВА ДЖЕИ ДИ [12]	32.3	27.4	61.9	33.7
G-RMI [11]	30.6	19.3	58.7	24.3

Машинные результаты по семантической сегментации (таблица 4: игнорируются классы). Методы с лучшим средним значением IOU также показывают лучшие результаты PQ. Обратите внимание, что G-RMI имеет довольно низкий PQ. Мы обнаружили, что это происходит потому, что он галлюцинирует множеством небольших участков классов, отсутствующих в изображении. Хотя это лишь незначительно влияет на IOU, который подсчитывает ошибки пикселей, это серьезно ухудшает PQ, который подсчитывает ошибки экземпляра.

напрямую вычисляем PQ. В таблице 4 мы сравниваем среднее значение

IoU, стандартный показатель для этой задачи, с PQ. Для городских пейзажей разрыв PQ между методами соответствует разрыву IoU. Для ADE20k разрыв намного больше. Это связано с тем, что в то время как IoU подсчитывает правильно предсказанный пиксель, PQ работает на уровне экземпляров. Смотрите таблицу 4 подпись для получения подробной информации. Паноптическая сегментация. Чтобы создать выходные данные алгоритма для PS, мы начинаем с неперекрывающихся сегментов экземпляра из процедуры, подобной NMS, описанной ранее. Затем мы объединяем эти сегменты с результатами семантической сегментации, разрешая любое перекрытие между классами thing и stuff в пользу класса thing (т. Е. Пикселю с меткой thing и stuff присваивается метка thing и ее экземпляр id). Эта эвристика несовершенна, но достаточна в качестве базовой.

Таблица 5 сравнивает PQ<sup>ст</sup> и PQ<sup>н</sup> вычисленный на com-

привязанные ("паноптические") результаты к производительности, достигнутой на основе отдельных прогнозов, рассмотренных выше. Для получения этих результатов мы используем работы-победители каждого соответствующего конкурса как для примера, так и для семантических задач. Поскольку перекрытия разрешаются в пользу вещей, PQ<sup>н</sup> является постоянным, в то время как PQ<sup>ст</sup> немного ниже для паноптических прогнозов. Визуализации паноптических выходов показаны на рисунке 9. Паноптическая сегментация "Человек против машины". Чтобы сравнить PQ человека и машины, мы используем машинные паноптические прогнозы, описанные выше. Для получения результатов для людей мы используем изображения с двойными комментариями, описанные в 6 и используйте bootstrap- ring для получения доверительных интервалов, поскольку эти наборы изображений невелики. Эти сравнения несовершенны, поскольку они используют разные тестовые изображения и усредняются по разным классам (некоторые классы, не имеющие совпадений в наборах тестов с двойной аннотацией, опущены), но они все равно могут дать некоторый полезный сигнал.

Мы представляем сравнение в таблице 6. Например, ма- китайцы лишь незначительно отстают от людей. С другой стороны, ма- китайский RQ значительно ниже человеческого, особенно на ADE20k и Vistas. Это подразумевает, что распознавание, то есть классификация, является основной проблемой для существующих методов. В целом, существует значительный разрыв между производительностью человека и машины. Мы надеемся, что этот разрыв вдохновит на будущие исследования для решения предлагаемой задачи panoptic segmentation.



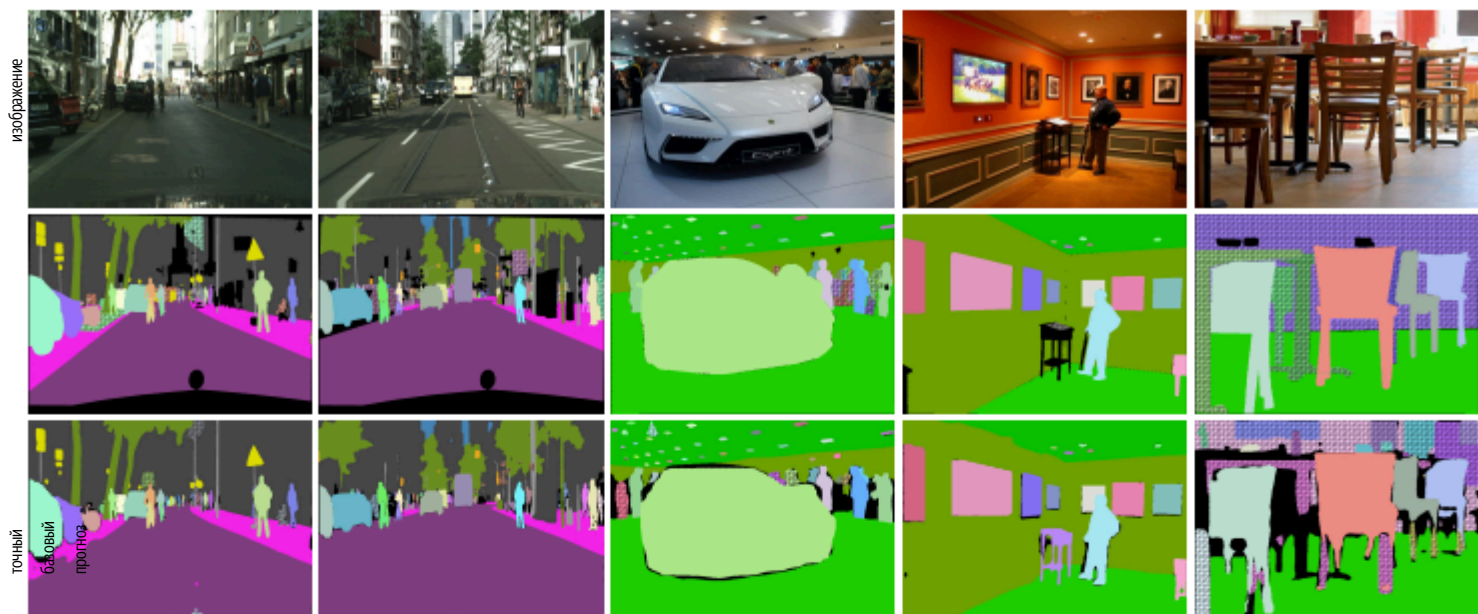


Рисунок 9: Результаты паноптической сегментации городских пейзажей (два слева) и ADE20k (три справа). Прогнозы основаны на объединенных выходных данных самых современных алгоритмов экземпляра и семантической сегментации (см. Таблицы 3 и 4). Цвета для сопоставленных сегментов ( $IoU > 0.5$ ) совпадают (рисунок перекрестной штриховки указывает на не совпадающие области, а черный - на немаркированные области). Лучше всего просматривается в цвете и с увеличением.

Городские пейзажи	PQ <sub>ex</sub>	PQ <sub>st</sub>	PQ <sub>h</sub>
машина-отдельная	61.2	66.6	54.0
машина-паноптик	PQ	66.4	54.0
ADE20k	n/a	ПК <sub>st</sub>	PQ <sub>h</sub>
машина-отдельная		27.4	41.1
машина-паноптик		24.5	41.1
Перспективы	PQ	ПК <sub>st</sub>	PQ <sub>h</sub>
машина-отдельная	n/a	43.7	35.7
машина-паноптик	38.3	41.8	35.7

Таблица 5: Паноптические и независимые прогнозы. "Машина"-

отдельные строки показывают PQ методов семантической сегментации и сегментации экземпляров, вычисляемых независимо (см. Также Таблицы 3 и 4).

Для "машинного паноптикума" мы объединяем непересекающиеся вещи и прочие предсказания, полученные с помощью самых современных методов, в истинную паноптическую сегментацию изображения. Из-за используемой эвристики слияния, PQ остается неизменным, пока PQ не слегка деградирует.

## 8. Будущее паноптической сегментации

Наша цель - продвигать исследования в новых направлениях, приглашая сообщество изучить новую задачу паноптической сегментации. Мы считаем, что предлагаемая задача может привести к ожидаемым и неожиданным инновациям. В заключение мы обсуждаем некоторые из этих возможностей и наши планы на будущее.

Мотивировано простотой, "алгоритм"

PS в этой статье основан на эвристике комбинация выходных данных от самых эффективных систем экземпляра и семантической сегментации.

Этот подход является базовым первым шагом, но мы ожидаем, что будет введено больше взаимосвязанных алгоритмов. В частности, мы надеемся, что PS будет стимулировать инновации по крайней мере в двух областях: (1) Глубоко интегрированные комплексные модели, которые одновременно решают

Городские пейзажи	PQ	площадь	PQ	PQ <sub>st</sub>	PQ <sub>h</sub>
пейзажи	69.6 <sup>+2.2</sup>	84.1 <sup>+0.8</sup>	82.0 <sup>+2.2</sup>	71.2 <sup>+2.3</sup>	67.4 <sup>+4.6</sup>
человек	61.2	480.9	74.4	66.4	54.0
машина	PQ	кв.	RQ	ПК <sub>st</sub>	PQ <sub>h</sub>
ADE20k человек	67.6 <sup>+2.0</sup>	85.7 <sup>+0.6</sup>	78.6 <sup>+2.1</sup>	71.0 <sup>+2.7</sup>	66.4 <sup>+4.2</sup>
машина	35.6	74.4	43.2	24.5	41.1
Перспективы	PQ	кв.	RQ	ПК <sub>st</sub>	PQ <sub>h</sub>
человек	57.7 <sup>+2.0</sup>	79.7 <sup>+0.7</sup>	71.6 <sup>+2.3</sup>	62.7 <sup>+2.8</sup>	53.6 <sup>+2.7</sup>
машина	38.3	73.6	47.7	41.8	35.7

Таблица 6: Производительность человека и машины. По каждому из-

согласованность сторонних наборов данных с человеком намного выше, чем производительность машины (приблизительное сравнение, подробнее см. Текст). Это особенно верно для RQ, в то время как SQ ближе. Разрыв самый большой на ADE20k и самый маленький на Cityscapes. Обратите внимание, что, поскольку доступен только небольшой набор пользовательских аннотаций, мы используем начальную загрузку и показываем диапазоны ошибок в 5-и 95-процентилях для человеческих результатов.

двойственная природа PS. Ряд подходов к сегментации экземпляров, включая [28, 2, 3, 18] предназначены для получения непересекающихся прогнозов экземпляров и могли бы служить основой такой системы. (2) Поскольку PS не может иметь перекрывающихся сегментов, некоторая форма "рассуждений" более высокого уровня может быть полезной, например, на основе расширения NMS, доступных для изучения [7, 16] к PS. Мы надеемся, что задача panoptic segmentation активизирует исследования в этих областях, что приведет к новым захватывающим прорывам в области видения.

Наконец, мы отмечаем, что задача panoptic segmentation была представлена как сложный трек обеими KOCO [25] и Картографические виды [35] проблемы с признанием и что предлагаемая задача уже начала набирать обороты в сообществе (например, [23, 48, 49, 27, 22, 21, 17] адрес PS).

## Ссылки

- [1] Э. Х. Адельсон. О видении материала: восприятие материалов Ху-человек и машины. В журнале Human Vision and Electronic Imaging, 2001. [1](#)
- [2] А. Арнаб и П. Х. Торр. Пиксельная сегментация экземпляра с помощью динамически создаваемая сеть. В [CVPR](#), 2017. [2](#), [3](#), [9](#)
- [3] м. Бай и Р. Urtasun. Глубокий водораздел трансформировать например Seg-мышления. В [CVPR](#), 2017. [3](#), [9](#)
- [4] Х. Цезарь, Дж. Уиджлингс и В. Феррари. COCO-Stuff: вещи и прочее классы в контексте. В CVPR, 2018. [2](#), [5](#)
- [5] Л.-К. Чен, Г. Папандреу, И. Коккинос, К. Мерфи и А. Л. Юилле. DeepLab: семантическая сегментация изображения с глубоким COBPa-lutional сети, Атрус свертки, и полностью подключен Ирк. PAMI, 2018. [1](#)
- [6] М. Кордтс, М. Омран, С. Рамос, Т. Рехфельд, М. Энцвейлер, Р. Бененсон, У. Франке, С. Рот и Б. Шиле. Набор данных cityscapes для понимания семантической городской сцены. В CVPR, 2016. [2](#), [3](#), [5](#), [6](#)
- [7] К. Десаи, Д. Раманан и К. К. Фаулкс. Дискриминационные модели для компоновки объектов нескольких классов. IJCV, 2011. [9](#)
- [8] П. Доллар, К. Воек, Б. Шиле и П. Перона. Обнаружение пешеходов: Оценка состояния техники. PAMI, 2012. [2](#)
- [9] М. Эверингем, С. А. Эслами, Л. Ван Гул, К. К. Уильямс, Дж. Винн, и А. Зиссерман. Проблема классов визуальных объектов PASCAL: ретроспектива. IJCV, 2015. [2](#)
- [10] А. Фатхи, Н. Канадзава и К. Мерфи. Испытание мест 2017: в-сегментация позиций, команда G-RMI. 2017. [8](#)
- [11] А. Фатхи, К. Янг и К. Мерфи. Испытание мест 2017: сцена разбор, команда G-RMI. 2017. [8](#)
- [12] Дж. Фу, Дж. Лю, Л. Го, Х. Тянь, Ф. Лю, Х. Лу, Ю. Ли, Ю. Бао и В. Янь. Places challenge 2017: разбор сцены, команда CASIA IVA JD. 2017. [8](#)
- [13] Б. Харихаран, П. Арбелаез, Р. Гиршик и Дж. Малик. Одновременный обнаружение и сегментация. В ECCV, 2014. [4](#), [5](#)
- [14] К. Хе, Г. Гкиоксари, П. Доллар, и Р. Гиршик. Маска R-CNN. В ICCV, 2017. [1](#), [2](#), [8](#)
- [15] Дж. Хосанг, Р. Бененсон, П. Доллар и Б. Шиле. Что делает для эффективные предложения по обнаружению? PAMI, 2015. [1](#)
- [16] Дж. Хосанг, Р. Бененсон и Б. Шиле. Изучение не максимального подавлению. PAMI, 2017. [9](#)
- [17] А. Кириллов, Р. Гиршик, К. Хе и П. Доллар. Паноптическая характеристика руга-сети mid. В CVPR, 2019. [9](#)
- [18] А. Кириллов, Е. Левинков, Б. Андрес, Б. Савчинский и К. Потер. InstanceCut: от ребер к экземплярам с помощью multicut. В CVPR, 2017. [2](#), [3](#), [9](#)
- [19] И. Коккинос. UberNet: обучение универсальной сверточной нейронной сети-работа для зрения низкого, среднего и высокого уровня с использованием разнообразных наборов данных и ограниченной памяти. В CVPR, 2017. [3](#)
- [20] А. Крижевский, И. Судцкевер и Г. Хинтон. Классификация ImageNet с глубокими сверточными нейронными сетями. В NIPS, 2012. [2](#)
- [21] Дж. Ли, А. Равентос, А. Бхаргава, Т. Тагава и А. Гайдун. Учимся соединять вещи и прочее. arXiv: 1812.01192, 2018. [9](#)
- [22] Q. Ли, А. Арнаб и П. Х. Торр. Панорама со слабым и полурегулируемым контролем-сегментация. В ECCV, 2018. [9](#)
- [23] Ю. Ли, Х. Чен, З. Чжу, Л. Се, Г. Хуан, Д. Ду и Х. Ван. Ориентированная на внимание объединенная сеть для паноптической сегментации. tation. arXiv: 1812.03904, 2018. [9](#)
- [24] Ю. Ли, Х. Ци, Дж. Дай, Х. Цзи и Ю. Вэй. Полностью сверточный экземпляр-осознанная семантическая сегментация. В CVPR, 2017. [2](#)
- [25] Т.-Я. Лин, М. Мэйр, С. Белони, Дж. Хейс, П. Перона, Д. Раманан, П. Доллар. Microsoft COCO: общие объекты в контексте. В ECCV, 2014. [2](#), [5](#), [9](#)
- [26] К. Лю, Дж. Юнь и А. Торральба. Поток ПРОСЕИВАНИЯ: плотное соответствие во всех сценах и их приложениях. PAMI, 2011. [2](#)
- [27] Х. Лю, К. Пэн, К. Ю, Дж. Ван, Х. Лю, Г. Ю и В. Цзян. Ань-сквозная сеть для паноптической сегментации. 2019. [9](#) arXiv: 1903.05027 ,
- [28] С. Лю, Дж. Цзя, С. Фидлер и Р. Уртасун. SGN: последовательная группировка сети, например, для сегментации. В CVPR, 2017. [2](#), [3](#), [9](#)
- [29] С. Лю, Л. Ци, Х. Цинь, Дж. Ши и Дж. Цзя. 17 ИЮНЯ: сегмент insatnce-задание на внедрение, команда победителей UCenter. 2017. [8](#)
- [30] Дж. Лонг, Э. Шелхамер и Т. Даррелл. Полностью сверточные сети. для семантической сегментации. В CVPR, 2015. [1](#), [2](#), [5](#)
- [31] Р. Ло, Б. Цзян, Т. Сяо, К. Пэн, Ю. Цзян, З. Ли, Х. Чжан, Г. Ю, Ю. Му и Дж. Сан. Places challenge 2017: сегментация инстансов, команда Megvii (Face++). 2017. [8](#)
- [32] Дж. Малик, П. Арбелаез, Дж. Каррейра, К. Фрагкиадаки, Р. Гиршик, Г. Гкиоксари, С. Гупта, Б. Харихаран, А. Кар и С. Тулсиани. Три "Р" компьютерного зрения: распознавание, реконструкция и реорганизованность. организация. PRL, 2016. [3](#)
- [33] Д. Р. Мартин, К. К. Фаулкс и Дж. Малик. Учимся распознавать природу-границы изображения гаl с использованием локальных сигналов яркости, цвета и текстуры. PAMI, 2004. [5](#)
- [34] И. Мисра, А. Шривастава, А. Гупта и М. Хеберт. Сетка для вышивания крестиком.-работает для многозадачного обучения. В CVPR, 2016. [3](#)
- [35] Г. Нойхольд, Т. Оплманн, С. Рота Було, и П. Концедер. The набор данных mapillary vistas для семантического понимания уличных сцен. В CVPR, 2017. [2](#), [3](#), [5](#), [6](#), [9](#)
- [36] П. О. Пиньейру, Р. Коллобер и П. Доллар. Учимся сегментировать объекты-кандидаты. В NIPS, 2015. [2](#)
- [37] С. Рен, К. Хе, Р. Гиршик и Дж. Сан. Быстрее R-CNN: к реальности-обнаружение объектов во времени с помощью сетей предложений регионов. В NIPS, 2015. [1](#)
- [38] О. Руссаковский, Дж. Денг, Х. Су, Дж. Краузе, С. Сатиш, С. Ма, З. Хуан, А. Карпати, А. Хосла, М. Бернштейн, А. К. Берг и Л. Фей-Фей. Масштабная задача визуального распознавания ImageNet. IJCV, 2015. [2](#)
- [39] Дж. Шоттон, Дж. Уинн, К. Потер и А. Криминизи. Textonboost: Совместная работа моделирование внешнего вида, формы и контекста для распознавания объектов нескольких классов. и segm. В ECCV, 2006. [2](#)
- [40] М. Сун, Б. Ким, П. Кохли и С. Саварезе. Связанные вещи и прочее посредством взаимодействия свойств объекта. PAMI, 2014. [3](#), [4](#)
- [41] Дж. Тиге и С. Лазебник. Поиск объектов: анализ изображений с использованием регионов и детекторы для каждого экземпляра. В CVPR, 2013. [3](#), [4](#)
- [42] Дж. Тиге, М. Найтхаммер и С. Лазебник. Разбор сцены с объектом примеры и порядок оклюзии. В CVPR, 2014. [2](#), [3](#), [4](#)
- [43] З. Ту, Х. Чен, А. Л. Юилле и С.-К. Чжу. Анализ изображений: объединение. сегментация, обнаружение и распознавание. IJCV, 2005. [2](#), [3](#)
- [44] Р. Вайан, К. Монрок и Ю. Лекун. Оригинальный подход к локализация объектов на изображениях. IEEE Pтос. по видению, имиджу и Обработка сигналов, 1994. [2](#)
- [45] С. Van Rijsbergen. Поиск информации. Лондон: Баттеруортс, 1979. [5](#), [7](#)
- [46] П. Виола и М. Джонс. Быстрое обнаружение объектов с использованием расширенного cas-набор простых функций. В CVPR, 2001. [2](#)
- [47] Д. Б. Уэст. Введение в теорию графов, том 2. Прентис Холл Река Анлер-Седло, 2001. [7](#)
- [48] Ю. Сюн, Р. Ляо, Х. Чжао, Р. Ху, М. Бай, Э. Юмер и Р. Уртасун. UPSNet: единая паноптическая сеть сегментации. arXiv: 1901.03784, 2019. [9](#)
- [49] Т.-Дж. Янг, М. Д. Коллинз, Ю. Чжу, Дж.-Дж. Хван, Т. Лю, Х. Чжан, В. Зе, Г. Папандреу и Л.-К. Чен. DeeperLab: анализатор изображений с одним снимком. arXiv: 1902.05093, 2019. [9](#)
- [50] Ю. Янг, С. Холлман, Д. Раманан и К. К. Фаулкс. Многослойный объектные модели для сегментации изображений. PAMI, 2012. [4](#)
- [51] Дж. Яо, С. Фидлер и Р. Уртасун. Описание сцены в целом: Совместное обнаружение объекта, классификация сцены и семантическая сегментация. В CVPR, 2012. [2](#), [3](#), [4](#)
- [52] Ф. Ю. и В. Колтун. Многомасштабная агрегация контекста с помощью расширенного контекста-публикации. В ICLR, 2016. [1](#)
- [53] Ю. Чжан, Х. Чжао и Дж. Ши. KONEЦ 17-го: семантическая сегментация. задание, команда-победитель PSPNet. 2017. [8](#)
- [54] Х. Чжао, Дж. Ши, Х. Ци, Х. Ван и Дж. Цзя. Разбор сцены пирамиды. сеть. В CVPR, 2017. [8](#)
- [55] Б. Чжоу, Х. Чжао, Х. Пуч, С. Фидлер, А. Барриузо и А. Торральба. Анализ сцены с помощью набора данных ADE20K. В CVPR, 2017. [2](#), [3](#), [5](#), [6](#)
- [56] Ю. Чжу, Ю. Тянь, Д. Мексатас и П. Доллар. Семантический амодальный сегмент-мышление. В [CVPR](#), 2017. [3](#)