

Научная статья
УДК 004.8:623

Актуальные подходы к решению задачи разметки данных, аугментации и синтеза предобученных моделей

Алексей Владимирович Ермоленко, Вячеслав Михайлович Полушкин,
Александр Сергеевич Горский

Аннотация. Статья посвящена исследованию актуальных подходов автоматической разметки снимков, получаемых с беспилотных летательных аппаратов, инструментарием современных топологий нейронных сетей, в том числе энкодеров и сверточных сетей. Рассмотрены аспекты формирования синтетических обучающих выборок. Предложены механизмы сохранения релевантной обучающей выборки посредством дополнительной нейросетевой коррекции коэффициентов геометрических параметров разметки.

Ключевые слова: сверточные нейронные сети; энкодеры; декодеры; беспилотные летательные аппараты; аугментация; синтетические данные; автоматическая разметка; функция потерь

Для цитирования: Ермоленко А.В., Полушкин В.М., Горский А.С. Актуальные подходы к решению задачи разметки данных, аугментации и синтеза предобученных моделей // Вооружение и экономика. 2024. №4(70). С. 36-42.

Original article

Current Approaches to the Problem Solution of Images Markup, Augmentation and Pre-trained Models Synthesis

Aleksei V. Ermolenko, Viacheslav M. Polushkin, Aleksandr S. Gorskii

Abstract. The article is devoted to the study of current approaches to automatic images marking obtained from unmanned aerial vehicles (UAV) by current neural network topologies including encoders and convolutional neural networks (CNN). The aspects of synthetic training samples creation are considered. The options for the relevant learning bases supporting by neural network correction of geometric image markup parameters coefficients are proposed.

Keywords: convolutional neural networks; encoders; decoders; unmanned aerial vehicles; augmentation; synthetic data; automatic marking; loss-function

For citation: Ermolenko A.V., Polushkin V.M., Gorskii A.S. Current Approaches to the Problem Solution of Images Markup, Augmentation and Pre-trained Models Synthesis. Vooruzhenie i ekonomika = Armament and Economics. 2024;70(4): 36-42. (In Russ.).

Введение

В настоящее время актуальным направлением является автоматическая высокоскоростная обработка изображений, которая невозможна без предварительной разметки соответствующих данных для обучения. Ручная разметка изображений, особенно аэрофото-снимков, – рутинный и дорогостоящий процесс, требующий значительных временных ресурсов и подготовленных специалистов. Актуальной задачей является автоматизация разметки. Одним из подходов к ее решению является применение нейронных сетей совместно с алгоритмическими методами предобработки снимков и аугментацией исходных данных.

1 Современные подходы к решению задачи автоматической разметки

Качество автоматической разметки снимков неструктурированных растровых изображений коррелировано с детерминацией искомых объектов в возможно точных контурах. Важным инструментом детерминации объектов на снимках является семантическая сегментация, представляющая собой высокоуровневую задачу обработки изображений. С ее помощью возможно выделить отдельные группы пикселей, соответствующие искомому классу, и попиксельно отнести их к конкретной области. Отметим, что в снимках со сложной фоновой обстановкой, объекты могут состоять из существенно различающихся по фотометрическим характеристикам элементов и иметь значительный разброс интенсивности

пикселей внутри одного класса. Оптимизировать решение таких задач позволяет семантическая сегментация, которая в отличие от классической сегментации не просто объединяет области по принципу цветового, текстурного или геометрического сходства, но и уточняет границы объектов с учетом семантического анализа.

Для задачи семантической сегментации сформировалось большое количество методов решения [1]. На современном этапе развития наибольшую эффективность демонстрируют методы, основанные на применении ансамблей сверточных нейронных сетей в комбинации с различными подходами к предобработке и постобработке изображений [2].

Применение ансамблей сверточных нейронных сетей для автоматической разметки снимков на основе семантической сегментации включает в себя основные этапы, представленные на рисунке 1. Подобная схема применима для большого массива накопленной фото (видео) информации, позволяющего сформировать репрезентативную обучающую выборку для корректного обучения нейронной сети.

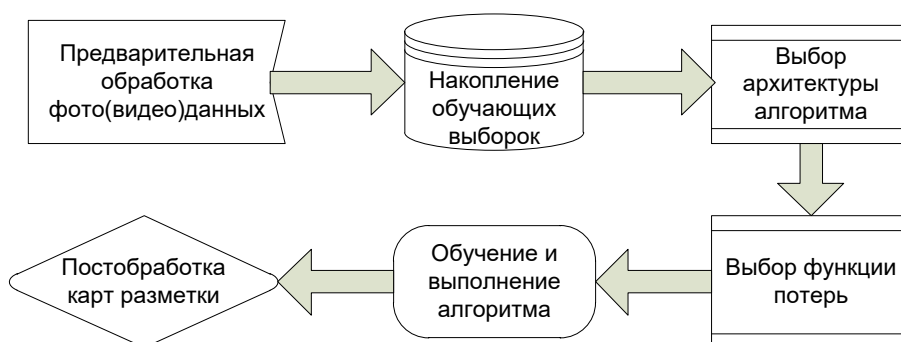


Рисунок 1 – Автоматизированная нейросетевая разметка изображений

2 Предобработка дата-сета графических данных

Успешная работа нейронной сети зависит от качественного наполнения дата-сета, в том числе от корректности определения линейных размеров объектов на проекции. В этой связи для снимков с беспилотных летательных аппаратов (БПЛА) важны сведения о высоте аэрофотосъемки, угловых координатах подвеса камеры летательного аппарата, фокусном расстоянии камеры и угле обзора. Отсутствие корректировки разметки приводит к значительному её смещению из-за угла съемки (особенно для высоких объектов), что важно в условиях выполнения задач, особенно в современной городской застройке и горной местности. Без коррекции разметки нейронная сеть часто обучается на нечто среднее, а в некоторых случаях, в принципе, рефлексировать не на искомый класс объектов.

Дополнительного повышения качества автоматической разметки, уточнения контуров детерминированных объектов и значительного снижения алгоритмической сложности возможно добиться применением дифференциально-разностных фильтров [3]. Фильтры позволяют выделить контура искомых объектов и установить коэффициенты смещения фрагментов снимков (видеоряда) с определением коэффициентов корреляции между ребрами опорных фигур. Если угол наклона камеры и высота съемки априорно известны, задачу можно упростить применением геометрического аппарата для восстановления тригонометрического соотношения ребер объектов. В некоторых случаях сдвиг векторной разметки снимка можно произвести силами оператора вручную. Однако применение такого подхода в условиях скоростной съемки с БПЛА практически невозможно по временным ограничениям.

В этой связи представляется целесообразным решение дополнительной подзадачи коррекции разметки посредством применения сверточных нейронных сетей для получения коэффициентов коррекции по вертикали и по горизонтали (рисунок 2).



Рисунок 2 – Разметка графических снимков с коррекций проекции сверточными нейронными сетями

3 Перспективные нейросети для автоматической разметки и семантической сегментации

Эффективными к применению в робототехнических комплексах (РТК) и БПЛА представляются сети U-Net. Данные топологии нейронных сетей демонстрируют хорошую устойчивость к помехам в изображении и сложной фоноцелевой обстановки. Эффективность применения U-Net для семантической сегментации и авторазметки предопределена тем, что сеть изначально разрабатывалась для анализа медицинских изображений постоянного ракурса и масштаба [4]. Сеть не требует большого количества изображений для обучения, так как имеет сравнительно небольшое количество параметров. U-Net использует «сквозные соединения» для сохранения пространственной информации. Карты признаков из сети-шифровальщика напрямую передаются и конкатенируются с картами признаков на соответствующих слоях сети-дешифровальщика параллельно с обычными сверточными слоями.

Сеть возможно реализовать в Python посредством нейросетевых модулей Keras и Theano, а впоследствии перенести на языки программирования низкого уровня (Си, Ассемблер) для реализации в микро-ЭВМ непосредственно в бортовых информационно-управляющих устройствах (БИУС) БПЛА (РТК).

Применимы различные модификации сети U-Net. Важным качеством U-net является возможность увеличения числа сверточных слоев. Для уменьшения оверфиттинга могут добавляться слои BatchNormalization и Dropout. В качестве целевой функции целесообразно использовать индекс Жаккара:

$$J = \frac{TP}{TP+FP+FN} = \frac{A \cap B}{A \cup B} = \frac{A \cap B}{A+B-A \cap B}, J \in (0,1), \tag{1}$$

где T – true, F – false, P – positive, N – negative.

Оптимальное решение разметки производится по сетке оптимальных параметров. При этом подбираются модели с максимальным счетом по результатам работы тестовой выборки. Применение сети в автоматическом режиме показывает высокие характеристики применимости до 10 типов целей.

Для уменьшения ложноположительных откликов сети эффективно введение индекса Тверски с уменьшающим коэффициентом:

$$S(X, Y) = \frac{|X \cap Y|}{|X \cap Y| + \alpha |X - Y| + \beta |Y - X|}, \tag{2}$$

где X – предсказанная разметка, Y – реальная разметка.

Удобным инструментом для тренировки синтезированных нейросетевых моделей представляется фреймворк TensorFlow, реализующий:

- развитый набор компонент для компоновки сетей;
- высокую скорость создания сетей для обучения в TensorFlow;
- возможность автоматической конвертации натренированной сети в C++ для переноса на нейрокомпьютеры и микро-ЭВМ.

4 Современные подходы к построению энкодеров и декодеров в задачах автоматической разметки

Архитектура U-Net состоит из двух соединённых между собой сетей: сети-шифровальщика (энкодера) для извлечения из изображения семантической информации в виде вектора признаков и сети-дешифровальщика (декодера) для превращения вектора признаков в матрицу нового изображения – маски классов. Чтобы сохранить пространственную информацию, карты признаков из энкодера с помощью сквозных соединений напрямую передаются в декодер и конкатенируются с картами признаков соответствующего разрешения декодера.

Для повышения ёмкости модели и точности работы сети в качестве сети-шифровальщика вместо исходного энкодера возможно применить сеть VGG-16. Альтернативными энкодерами являются ResNet34, InceptionV3, MobileNetV2, EfficientNetB0 [5]. Кроме того, может быть применена сеть SegNet [6] для операции рассоединения.

Новизна предлагаемых подходов в том, что при операции объединения по максимуму на этапе свертки в сети-шифровальщике индексы максимальных значений сохраняются и позже используются для дискретизации соответствующей карты признаков в сети-дешифровальщике. Для этого производится обратное рассоединение с использованием сохраненных индексов. При использовании сетей LinkNet [8] вместо конкатенации возможно применять сложение карт признаков.

Применимы и другие подвиды сетей. В частности, альтернативная архитектура DeepLab [9] обладает тремя качественными отличиями:

- свертка фильтрами с повышенной дискретизацией;
- пространственное пирамидальное объединение (ASPP) таких фильтров для сегментирования объектов в разных масштабах.
- повышение качества локализации границ объектов комбинированием методов глубоких сверточных нейронных сетей и вероятностных графических моделей (CRF) для учёта контекстной информации.

В CFNet [5] для учёта контекста возможно производить оценку вероятности совместного проявления различных признаков посредством модуля ACOFM. Приём объединения карт признаков, с помощью которого сеть-дешифровальщик получает информацию о глобальном контексте, используется в архитектуре PSPNet [6].

Повышение производительности отличным от PSPNet способом возможно посредством снижения потерь контекста в RefineNet [5]. В такой сети возможно итеративно объединить повышающие разрешение векторы признаков. Таким образом, возможно повысить разрешение разметки. При дополнительном применении модификации DANet [6] возможно применить механизм связей для моделирования зависимостей как внутри каждого канала, так и между каналами.

Модель Mask R-CNN представляет собой развитие методов детектирования и решает одновременно две задачи: строит ограничивающее окно разметки объекта и одновременно в этом прямоугольнике проводит сегментацию с уточнением границ. При этом R-CNN генерирует маску с разделением общей маски на связные области. Недостатком является фиксированное разрешение маски каждого объекта, то есть для сложных контуров детерминированных объектов граница заведомо получится упрощённой.

5 Функции потерь

Распространённой проблемой в области анализа изображений является детектирование или сегментирование очень маленького «аномального» региона большого изображения (например, выделение небольшого объекта на фоне большого количества зелёных насаждений). Легко классифицируемые примеры составляют большую часть обучающей выборки и доминируют при расчёте функции потерь. Сложные примеры, на которых сеть ошибается, практически игнорируются при обучении, так как их количество относительно мало. Проблему несбалансированности данных можно уменьшить изменением обучающей выборки либо введением функции потерь.

Одним из возможных подходов является кросс-энтропия. Кросс-энтропийная (CE) функция потерь часто используется в задачах автоматической разметки данных. Её выходной сигнал представляет собой значение вероятности в диапазоне от 0 до 1. Величина кросс-энтропийной функции потерь увеличивается, когда прогнозируемая вероятность отклоняется от целевой метки.

Альтернативную функцию потерь можно построить на основе коэффициента Дайса – Сёренсен, определяющего меру сходства между множествами. В некоторых случаях целесообразно изменить форму функции потерь таким образом, чтобы повысить приоритизацию на редко встречающихся примерах, для чего ввести модулирующий коэффициент $(1 - p_t)^\gamma$ к кросс-энтропийной функции потерь, где фокусирующий параметр $\gamma \geq 0$. В случае доминирования объектов одного из классов сеть при обучении будет пытаться устранить небольшие ошибки на объектах доминирующих классов, а существенные ошибки редких классов будет игнорировать. Когда объект классифицирован неправильно, и предсказанная вероятность p_i принадлежности объекта к классу i небольшая, модулирующий фактор близок к единице, и значение функции потерь не изменяется. При вероятности $p_i \rightarrow 1$ коэффициент модуляции обнуляется. Параметр фокусировки плавно регулирует скорость, с которой у «простых» примеров понижаются веса. Когда коэффициент является нулевым, фактическая функция потерь становится тождественна кросс-энтропийной.

6 Улучшение точности классификации (постобработка)

Для улучшения пространственной поддержки разметки необходимо учитывать зависимости между целевыми переменными, для чего представляется целесообразным использовать структурный подход на основе модели условных случайных полей (CRF) [7].

В отличие от дискретных классификаторов, определяющих метку для конкретного образца без учёта меток соседних объектов, CRF может учитывать контекст изучаемого снимка. Учёт контекста заключается в априорном допущении, что соседний к данному сегмент имеет тот же класс (то есть границы между классами относительно редки), а распределение классов на сцене подчинено закону нормального распределения. При этом вычислять константу нормализации не требуется. Графическая интерпретация данного математического допущения представляет собой условное случайное поле.

Успешное применение случайных полей сглаживает результаты разметки и устраняет ложные элементы.

7 Аугментация и синтез дата-сетов малого объема

Современным трендом разработчиков решений в области искусственного интеллекта является широкое применение аугментации и синтетических данных. Основной причиной применения таких инструментов является малый объем баз знаний обучающих выборок. Аугментация данных позволяет увеличить объем выборки для обучения модели путем управляемой модификации существующих данных [8]. Аугментация эффективна для мало-документированных объектов оперативного интереса, а также в условиях временных и ресурсных ограничений, иными словами, когда качественные и количественные характеристики дата-сета не отвечают минимальным требованиям алгоритмов обучения сети. Снимки одной и той же местности в динамике имеют существенные отличия. Они производятся в разное время суток, имеют различную цветопередачу, условия освещенности. На снимках могут присутствовать тени от облаков, задымленность и другие воздействующие факторы. Отображение подстилающей поверхности (цвет травы, асфальта, бетона) может сильно отличаться даже у изображений, сделанных в одном и том же месте одним и тем же БПЛА. Искусственная имитация таких фотометрических особенностей позволяет повысить обобщающие способности нейросети в процессе формирования обучающей выборки.

Механизмами аугментации могут быть изменения разрешения и перекрытие фрагментов. Использование масштабирования позволяет моделировать возможную разницу

в размерах объектов. Нарезка фрагментов с перекрытием позволяет использовать большее количество контекстов.

Наиболее применимые методы аугментации:

- операции с цветом (модификация цветового градиента, увеличение контрастности или яркости);
- геометрические операции (поворот изображения, отражение изображения и т.п.);
- операции с объектами (поворот объекта, отражение объекта и т.п.).

Методы оптимизированной аугментации данных увеличивают точность модели в метрике mAP (mean Average Precision) как для COCO-моделей, так и для PASCAL-VOC [8]. Практические результаты показывают, что качественная аугментация работает лучше, чем методы регуляризации для распознавания объектов.

Одновременно с этим, вызывает озабоченность стремление разработчиков к повсеместному наращиванию аугментированных и синтетических данных в обучающем дата-сете. При этом аугментирование нередко применяется для дообучения нейросети скрытым элементам классифицируемых объектов в условиях отсутствия достоверной модели объекта. Это несет риски ложных реакций синапсов нейросети, непрогнозируемому росту ошибок 1-го и 2-го рода.

Полученные эмпирическим путем результаты практического тестирования авторазметчиков, основанных на сверточных нейронных сетях и семантической аугментации, показывают, что доля аугментированных данных в общем объеме выборки не должна превышать 20-25%. По мере уточнения знаний о модели объекта оперативного интереса доля аугментированных и синтетических данных в общем дата-сете может повышаться.

Заключение

Анализ современных подходов демонстрирует многообразие возможностей автоматической разметки данных. Наиболее эффективным представляется синтез модифицированных сетей U-Net и VGG. Методы предобработки, в том числе контурного анализа, позволяют снизить инвариантность классификации и идентификации графических примитивов в ходе семантической сегментации.

Использование сверточных нейронных сетей в задачах автоматической разметки – наиболее эффективный на сегодня подход. Вместе с тем, перспективной альтернативой является использование генеративных состязательных сетей, работающих без прямого использования функции потерь.

Аугментация и синтезирование данных позволяет повысить объем обучающей выборки, однако не всегда увеличение базы знаний аугментированными данными гарантирует сохранение качества обучения нейронной сети.

В целом, синтез применения нейросетевых корректоров разметки, предобработки снимков дифференциально-разностными фильтрами, инструментов семантической сегментации, аугментации и синтезирования данных позволяет снизить алгоритмическую сложность автоматической разметки и повысить эффективность систем технического зрения БПЛА.

Список источников

1. Горбачёв В.А., Криворотов И.А., Маркелов А.О., Котлярова Е.В. Семантическая сегментация спутниковых снимков аэропортов с помощью сверточных нейронных сетей // Компьютерная оптика. 2020. Т.44. №4. С. 636-645.
2. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation. 2015 IEEE CVPR – Conference on Computer Vision and Pattern Recognition (USA, Boston, 2015, June 07-12). DOI: 10.1109/CVPR.2015.7298965.
3. Ермоленко А.В., Князев Р.И., Полушкин В.М. Детерминация примитива на кластеризованной сцене посредством синтеза матричных фильтров и градиентных алгоритмов обучения //

Радиолокация, навигация, связь: сб. трудов XXIX Междунар. науч.-техн. конф. (г. Воронеж, 2023, 18-20 апреля). В 5 т. Т.1. Воронеж: ВГУ, 2023. С. 287-295.

4. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention: 18th International Conf. (Germany, Munich, 2015, October 05-09). Part III. Springer, 2015. P. 234-241.

5. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. 2016 IEEE CVPR – Conference on Computer Vision and Pattern Recognition (USA, Las Vegas, 2016, June 27-30). DOI: 10.1109/CVPR.2016.90.

6. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(12): 2481-2495. DOI: 10.1109/TPAMI.2016.2644615.

7. Соловьев Р.А., Тельпухов Д.В., Кустов А.Г. Автоматическая сегментация спутниковых снимков на базе модифицированной сверточной нейронной сети UNET // Инженерный вестник Дона. 2017. №4(47). С. 75.

8. Валетов Т.Я. Применение открытых картографических сервисов (Google, Яндекс, OSM) при создании исторических ГИС: разработка цифровой карты Транссибирской магистрали // Историческая информатика. 2021. №3(37). С. 19-37.

Информация об авторах

А.В. Ермоленко – кандидат технических наук, SPIN код автора 8544-4195.

В.М. Полушкин – кандидат технических наук.

А.С. Горский – кандидат технических наук, SPIN код автора 2164-7791.