

Бигеев Амаль Денисович

студент, МГТУ им Н. Э. Баумана,

РФ, г. Москва

**ВЛИЯНИЕ АРХИТЕКТУРЫ ПИРАМИДЫ ПРИЗНАКОВ НА ТОЧНОСТЬ
РАСПОЗНАВАНИЯ МАЛОРАЗМЕРНЫХ ОБЪЕКТОВ
ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ**

В статье исследуется проблема снижения пространственного разрешения при обработке данных дистанционного зондирования Земли сверточными нейронными сетями. Объектом исследования выступают методы построения многомасштабных представлений изображений для задачи обнаружения объектов, площадь проекции которых не превышает 512 квадратных пикселей. Проведен математический анализ потери информации о малоразмерных объектах при прохождении тензора через сверточные слои с шагом дискретизации 8 и выше. Выполнено сравнение трех архитектур объединения признаков: нисходящей пирамидальной сети, сети агрегации путей и двунаправленной пирамидальной сети с обучаемыми весами слияния. Обоснована необходимость внедрения дополнительного уровня высокого разрешения с шагом дискретизации 4. Результаты вычислительных экспериментов на наборе данных DOTA показывают, что модификация двунаправленной пирамиды признаков указанным уровнем повышает точность обнаружения малоразмерных объектов до 0,321.

ABSTRACT

The article investigates the problem of spatial resolution reduction when processing aerospace images using convolutional neural networks. The object of the study is methods for constructing multiscale image representations for the task of localizing elements with a projection area not exceeding 512 square pixels. A mathematical analysis of information loss during tensor propagation through convolutional layers

with a discretization step of 8 and higher is conducted. A comparison of three feature fusion architectures is performed: a top-down feature pyramid network, a path aggregation network, and a bidirectional feature pyramid network with trainable fusion weights. The necessity of introducing an additional high-resolution level with a discretization step of 4 is substantiated. The results of computational experiments on the DOTA dataset show that modifying the bidirectional feature pyramid with the specified level increases localization accuracy to 0.321.

Ключевые слова: аэрокосмические изображения, многомасштабное представление, пирамида признаков, пространственное разрешение, шаг дискретизации, локализация элементов инфраструктуры, топология сети.

Keywords: aerospace images, multiscale representation, feature pyramid, spatial resolution, discretization step, infrastructure elements localization, network topology.

Введение

Обработка данных дистанционного зондирования Земли методами компьютерного зрения сопряжена с проблемой обнаружения объектов малой площади [4]. На изображениях с пространственным разрешением от 0.5 до 5 метров на пиксель объекты инфраструктуры занимают площадь от 4 до 512 квадратных пикселей [5].

Архитектуры сверточных нейронных сетей применяют операции субдискретизации для увеличения рецептивного поля и извлечения семантически сложных признаков. Данный процесс приводит к последовательному уменьшению пространственных размеров карт признаков. Для исходного изображения размером 512 на 512 пикселя глубокие слои сети формируют карты размером 32 на 32 пикселя. В результате объекты, занимавшие несколько пикселей на исходном изображении, исчезают на глубоких уровнях сети.

Для решения указанной проблемы применяются методы формирования многомасштабного представления — пирамиды признаков. Они объединяют

семантически насыщенные признаки глубоких слоев с пространственно точными признаками ранних слоев.

Цель работы — исследование влияния различных архитектур пирамид признаков на точность обнаружения малоразмерных объектов и обоснование оптимальной конфигурации уровней дискретизации.

Математическое обоснование потери пространственной информации

Процесс извлечения признаков базовой сверточной сетью сопровождается понижением разрешения. Шаг дискретизации определяет, во сколько раз пространственные размеры выходной карты признаков меньше размеров исходного изображения.

Стандартные реализации пирамид признаков используют уровни с шагами дискретизации 8, 16, 32, 64 и 128. Рассмотрим объект размером 4 на 4 пикселя на входном изображении. При проецировании данного объекта на карту признаков с минимальным стандартным шагом дискретизации 8, его размер вычисляется как отношение исходного размера к шагу дискретизации. Результат составляет 0,5 элемента карты признаков.

Объект, занимающий менее одного элемента карты признаков, не генерирует достаточного активационного сигнала для последующих слоев классификации и регрессии координат. Математически это означает потерю информации об объекте.

Для сохранения информации требуется введение в пирамиду признаков уровня с шагом дискретизации 4. При таком шаге объект размером 4 пикселя проецируется ровно в один элемент карты признаков, что делает возможным вычисление градиентов при обратном распространении ошибки.

Анализ архитектур объединения признаков

В ходе исследования рассмотрены три подхода к передаче информации между уровнями карт признаков разного разрешения.

Первый подход — нисходящая пирамидальная сеть [1]. Метод использует нисходящий путь для передачи семантической информации от глубоких слоев к мелким. Карта признаков низкого разрешения увеличивается в два раза методом билинейной интерполяции и поэлементно складывается с картой более высокого разрешения. Недостаток подхода заключается в равном весе складываемых карт, что не учитывает различную релевантность уровней для объектов разных масштабов.

Второй подход — сеть агрегации путей [2]. Метод расширяет нисходящую сеть добавлением восходящего пути. Восходящий путь передает пространственную информацию от мелких слоев к глубоким через операции свертки с шагом 2. Данная архитектура создает короткие соединения между уровнями, однако сохраняет проблему неадаптивного взвешивания при сложении тензоров.

Третий подход — двунаправленная пирамидальная сеть [3]. Метод вводит обучаемые скалярные веса для каждого входа в операции слияния. Сеть самостоятельно определяет вклад каждого уровня в итоговую карту признаков в процессе обучения. Для устранения отрицательных значений весов применяется функция активации линейной ректификации. Нормализация суммы весов выполняется делением каждого веса на общую сумму весов с добавлением константы численной стабильности. Топология сети упрощена за счет удаления узлов, имеющих только один входной признак.

Организация вычислительных экспериментов

Экспериментальная проверка проводилась на наборе данных DOTA [6]. Выборка содержит изображения с аннотациями объектов инфраструктуры. Для оценки точности использована подвыборка малоразмерных объектов, площадь которых строго меньше 512 квадратных пикселей.

В качестве базовой сети извлечения признаков применялась архитектура MobileNet. Обучение выполнялось с использованием алгоритма оптимизации SGD. Размер пакета данных составлял 16. Базовый шаг обучения установлен на уровне одной десятитысячной с применением косинусного убывания.

В ходе экспериментов обучались четыре конфигурации нейронной сети, отличающиеся структурой модуля слияния признаков:

1. Конфигурация с нисходящей пирамидальной сетью и стандартным набором уровней дискретизации от 8 до 128.
2. Конфигурация с сетью агрегации путей и набором уровней дискретизации от 8 до 128.
3. Конфигурация с двунаправленной пирамидальной сетью и набором уровней дискретизации от 8 до 128.
4. Конфигурация с двунаправленной пирамидальной сетью, расширенной дополнительным уровнем с шагом дискретизации 4.

Оценка результатов производилась по метрике средней точности обнаружения малоразмерных объектов при пороге пересечения предсказанной и истинной рамок, равном 0,5. Вычислительная сложность оценивалась по количеству обучаемых параметров модуля слияния признаков.

Результаты экспериментов

Результаты оценки точности и вычислительной сложности для четырех конфигураций представлены в таблице 1.

Таблица 1 — Сравнение методов построения пирамид признаков

Метод объединения признаков	Шаги дискретизации	Точность на малоразмерных объектах	Количество параметров модуля, млн
Нисходящая пирамида	8, 16, 32, 64, 128	0,294	4,42
Сеть агрегации путей	8, 16, 32, 64, 128	0,311	4,74
Двунаправленная пирамида	8, 16, 32, 64, 128	0,301	3,04

Двунаправленная пирамида	4, 8, 16, 32, 64	0,321	2,66
--------------------------	------------------	-------	------

Анализ данных таблицы 1 показывает, что переход от нисходящей пирамиды к сети агрегации путей дает прирост точности на 0,01, однако приводит к увеличению числа параметров на 2,08 млн. Применение двунаправленной пирамиды со стандартными уровнями дискретизации увеличивает точность до 0,301 при одновременном снижении числа параметров до 3,04 млн. Данный эффект объясняется наличием обучаемых весов слияния, которые адаптивно фильтруют нерелевантную информацию, и удалением избыточных узлов в топологии сети.

Заключение

Проведено исследование методов построения пирамид признаков для задачи распознавания малоразмерных объектов на данных дистанционного зондирования. Показано, что использование стандартных уровней пирамиды с минимальным шагом дискретизации 8 приводит к потере информации об объектах размером менее 8 пикселей.

Анализ архитектур слияния признаков показал преимущество двунаправленной пирамидальной сети с обучаемыми весами. Данный метод обеспечивает адаптивное объединение информации при минимальном количестве параметров.

Модификация двунаправленной пирамиды путем добавления уровня с шагом дискретизации 4 устраняет проблему потери пространственного разрешения для объектов площадью до 16 квадратных пикселей.

Результаты вычислительных экспериментов подтверждают эффективность предложенной конфигурации: точность обнаружения малоразмерных объектов возрастает с 0,294 до 0,321 при сохранении низкой вычислительной сложности модуля.

Список литературы:

1. Liu S. Path Aggregation Network for Instance Segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. C. 8759—8768.
2. Tan M., Pang R., Le Q. V. EfficientDet: Scalable and Efficient Object Detection // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. C. 10781—10790.
3. Chen L., Wang H., Zhou Y. BPD-YOLO: A Small Object Detection Method for Aerial Images Based on Improved YOLOv8 // Scientific Reports. 2025. T. 15. C. 16878.
4. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L. C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. C. 4510—4520.
5. Xia G.-S. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. C. 3974—3983.
6. Yu J. UnitBox: An Advanced Object Detection Network // Proceedings of the ACM International Conference on Multimedia. 2016. C. 516—520.