УДК 004.85:004.032.26

GraFeat-Merge: объединение моделей компьютерного зрения на основе графовых представлений

Дженжеруха К. А.[1], Гридусов Д. Д.[1]
Научный руководитель – к.т.н., доцент, зам. директора ВШЦК Романов А. А.
Научный консультант – к.т.н., доцент ВШЦК Азимов Р. Ш.[1]
[1]Университет ИТМО
kdzhenzherukha@gmail.com

Введение

В настоящее время модели машинного обучения в области компьютерного зрения используются для решения огромного количества задач – от классификации объектов до детекции семантически сложных аномалий или обеспечения точной реконструкции окружения для автономного транспорта. При решении любой из этих задач важно, чтобы сеть извлечения признаков из изображений (backbone) на выходе содержала наиболее информативное представление о том, что содержится на изображении. Предобученные модели за счет большого разнообразия данных обучающей выборки, часто могут извлекать качественные признаки, но для сложных сценариев или данных из специфичного домена они могут быть недостаточными или, наоборот, избыточными, а дообучать модели или обучать их с нуля достаточно затратно (даже при наличии GPU).

В рамках исследования были изучены методы объединения различных моделей машинного обучения, но подавляющее большинство из них применимы лишь к задачам обработки естественного языка (к примеру, MergePipe [1]) в силу унифицированного вида токенов, к которым преобразуется текст. Но в области компьютерного зрения в силу фундаментальных ограничений, таких как различия архитектур и видов выходов слоев моделей, эти методы неприменимы в случае, когда признаковые представления, извлекаемые разными моделями отличаются.

Основная часть

В данной работе разрабатывается способ слияния моделей разных архитектур (как предобученных на датасетах общего назначения, так и обученных на конкретных доменах) с помощью графов.

Предлагаемый подход состоит из нескольких этапов:

– на первом шаге для каждой модели с активации последнего сверточного слоя извлекаются признаки изображения – тензор размерности (B, C, H, W), где B – размер батча, C – количество каналов после сверточного слоя, (H, W) – размер изображения после прохождения сверточных слоев;

– далее формируется полносвязный гетерогенный граф, в котором вершинами выступают извлеченные вектора признаков (H x W векторов размерности C), а ребра делятся внутренние (между признаками одной модели) и внешние (между признаками текущей модели и всех остальных);

– на каждом шаге состояние вершин обновляется с помощью обучаемых слоев внутреннего и перекрестного внимания: таким образом обеспечивается обмен знаниями как внутри одной модели, так и между ними;

– после этого из графа извлекается один вектор [2], который служит финальным представлением изображения и может быть использован для решения конкретной задачи;

Подобный подход к динамическому обновлению состояний вершин графа и использованию его для решения задачи компьютерного зрения был успешно применен в алгоритме сопоставления ключевых точек [3], что дает обоснование данной работе.

Выводы

Первые эксперименты на датасете MNIST (базовый датасет для бенчмарков в области моделей компьютерного зрения) показывают, что предлагаемое решение по объединению моделей компьютерного зрения может успешно применяться в реальных пайплайнах. Комбинация признаковых представлений на выходе разных моделей (как по архитектуре в рамках одного типа, так и разных типов моделей) с помощью внутреннего и перекрестного внимания позволяет выделять наиболее значимые и информативные признаки, отражающие основные детали изображений и позволяющие получать наиболее точное финальное векторное представление.

Литература

1. Y. Wang, Y. Gu, Z. Wang, K. Li, Y. Yang, Z. Yan, C. Xie, J. Wu, H. Yang. MergePipe: A Budget-Aware Parameter Management System for Scalable LLM Merging. 2026 // arXiv:2602.13273. URL: https://arxiv.org/abs/2602.13273 (дата обращения: 18.02.2026 г.)
2. J. Wang, Y. Guo, L. Yang, Y. Wang. Enabling Homogeneous GNNs to Handle Heterogeneous Graphs via Relation Embedding. 2022. // IEEE Transactions on Big Data 9 (2023) 1697-1710
3. P.-E. Sarlin, D. DeTone, T. Malisiewicz, A. Rabinovich. SuperGlue: Learning Feature Matching with Graph Neural Networks. 2019. // CVPR 2020.