Интеграция графов знаний с компьютерным зрением: методы и подходы

Автор: Раконяц Даниела

Магистратура: «Методы и технологии искусственного интеллекта»

Руководитель: Кандидат физ.-мат. наук Куратов Ю.М.

Введение

- Современные задачи компьютерного зрения:
  - Навигация
  - Классификация изображений
  - Распознавание объектов

- Проблемы традиционных методов:
  - Сложности работы с зашумлёнными или сложными данными
  - Отсутствие контекстуального анализа

- Решение: использование графов знаний:
  - Структурированное представление связей между объектами
  - Улучшение семантической обработки данных

Цели и подходы

- Цель: интеграция графов знаний для повышения точности и эффективности моделей компьютерного зрения

- Три основных направления:

  1. Согласование графов знаний с визуальными данными для навигации

  2. Применение графов для классификации изображений

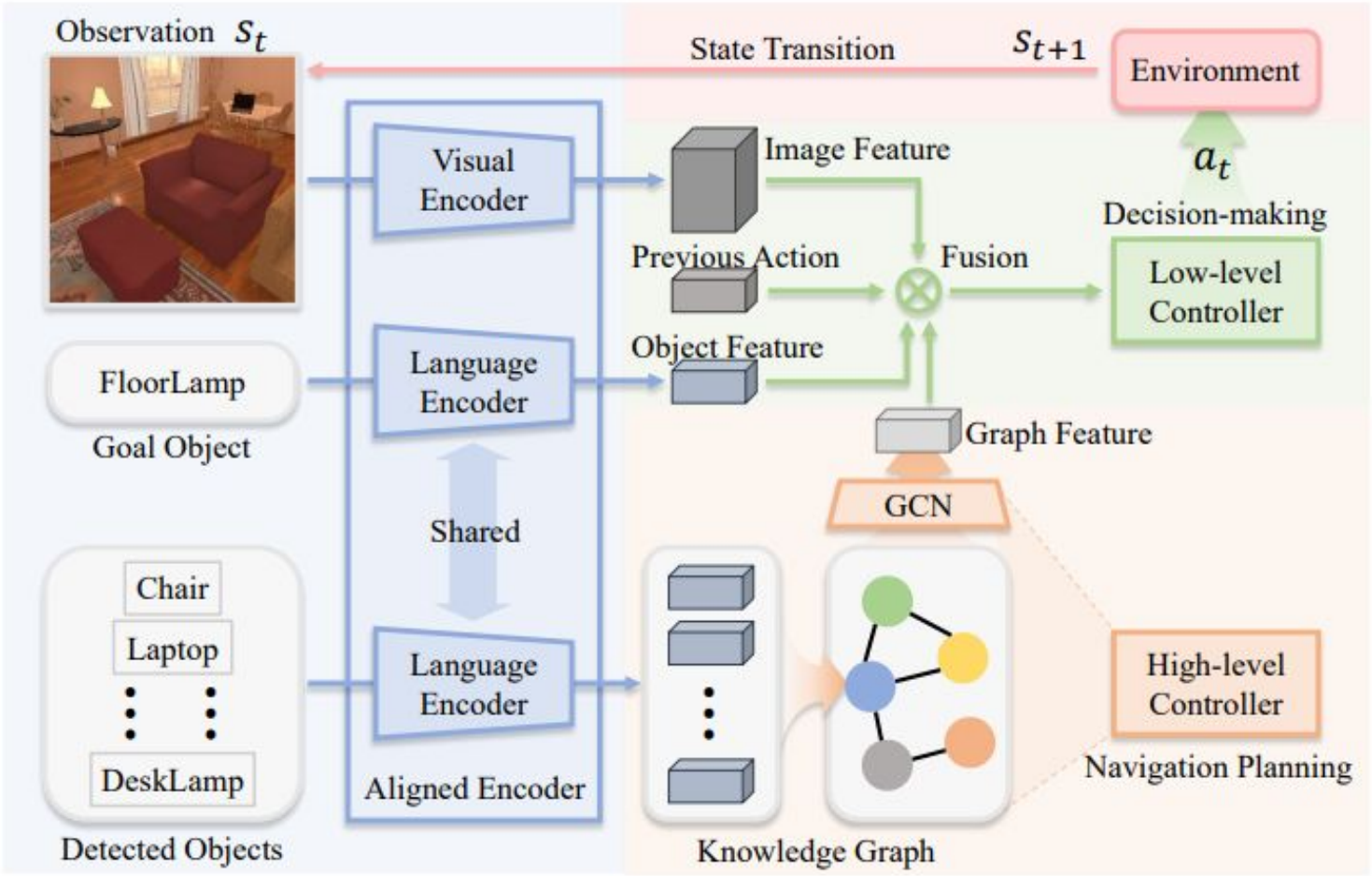
  3. Контекстное распознавание объектов

Метод 1 — Навигация на основе графов знаний (AKGVP)

Задача: навигация по объектам-целям (Object-goal Navigation)

Методология:

- Мультимодальная интеграция текстовых и визуальных данных
- Построение иерархической модели сцены
- Использование симулятора AI2-THOR для тестирования


Observation s_t
State Transition
s_{t+1}
Environment
a_t
Decision-making
Visual Encoder
Image Feature
Previous Action
Fusion
Low-level Controller
FloorLamp
Goal Object
Language Encoder
Object Feature
Graph Feature
GCN
Shared
Chair
Laptop
DeskLamp
Language Encoder
Aligned Encoder
Detected Objects
Knowledge Graph
High-level Controller
Navigation Planning

Метод 2 — Классификация изображений (GSNN)

Основной подход:

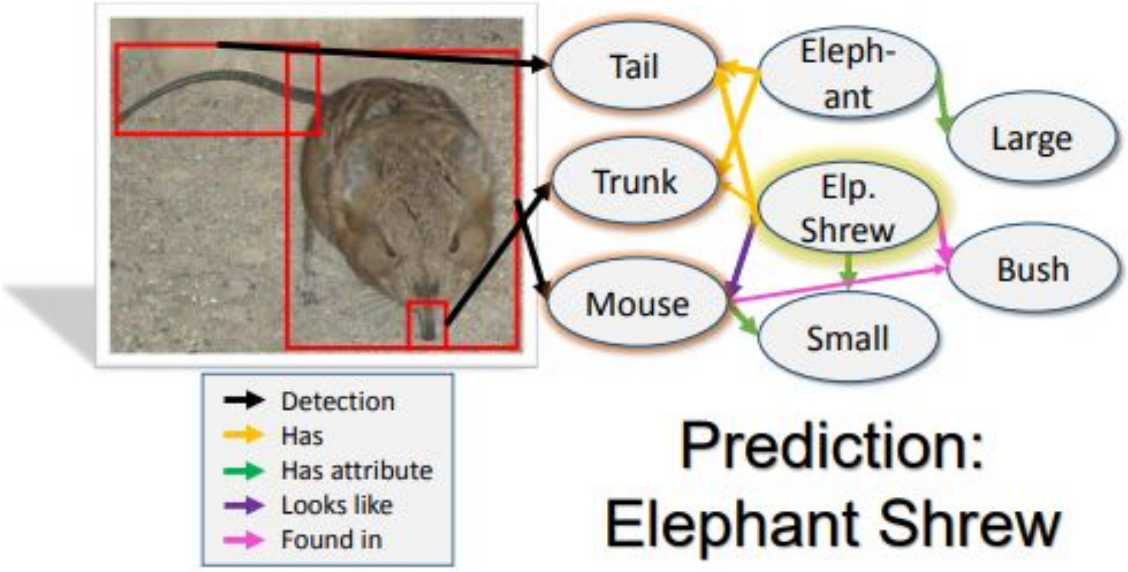
- Использование графовых нейронных сетей (GGNN) и их модификации GSNN
- Построение графов на основе Visual Genome и WordNet

Результаты:

- Повышение точности на наборах Visual Genome и COCO
- Пример: COCO, mAP вырос с 69.86% до 77.57%

Проблемы:

- Ограничения масштабируемости
- Зависимость от детекторов объектов



Пример того, как семантические знания о мире помогают в классификации. Здесь мы видим слоновую землеройку. Люди способны правильно классифицировать её на основе того, что мы знаем о слоновой землеройке и других похожих животных.

Метод 3 — Контекстное распознавание объектов

Типы контекста:

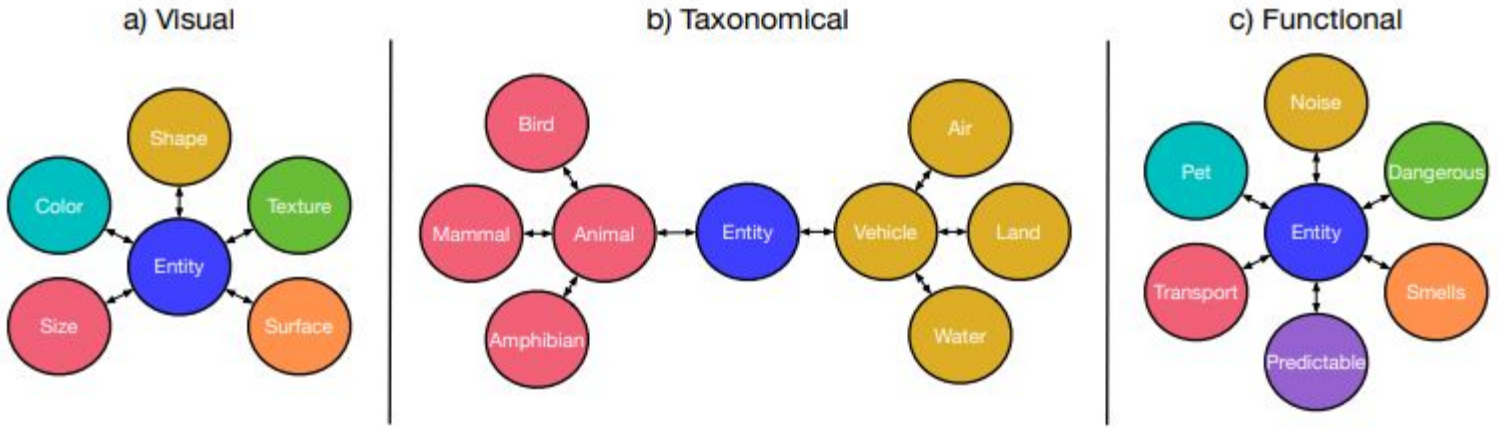
1. Визуальный: цвет, форма, текстура
2. Таксономический: связи между категориями
3. Функциональный: свойства и функции объектов

Методология:

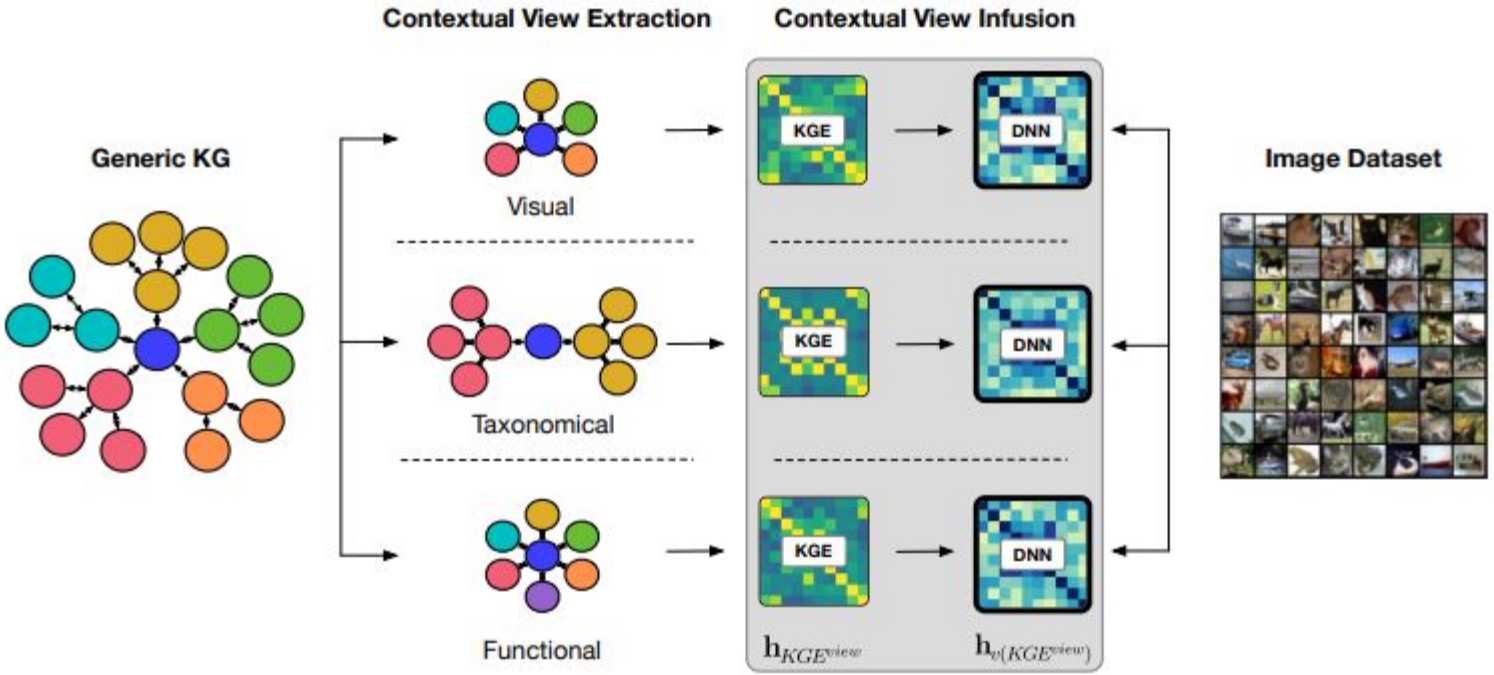
- Извлечение подграфов знаний
- Интеграция представлений графа в нейронные сети

Результаты:

- Улучшение точности в условиях смены домена (например, Mini-ImageNet)

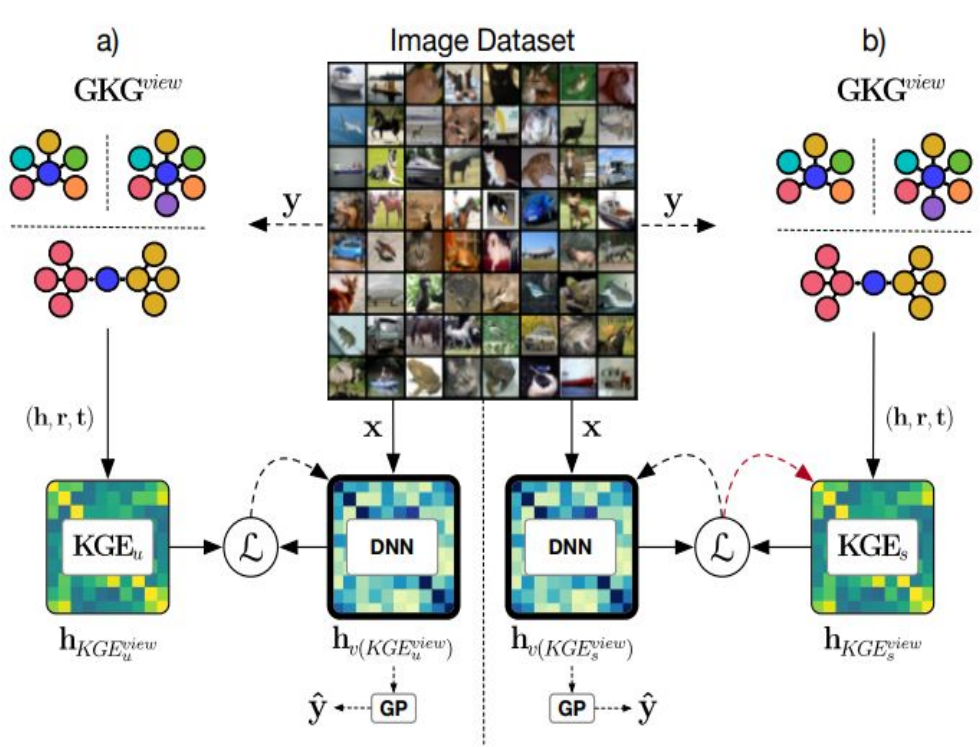


Контекст может проявляться по-разному. Основываясь на представлениях о том, как люди воспринимают мир, мы представляем три контекстных представления общего графа знаний, а именно визуальное, таксономическое и функциональное представления.



Подход к изучению контекстных представлений изображений состоит из двух основных частей: 1) извлечение контекстного представления; 2) внедрение контекстного представления.

Схема а	Схема б
Строится модель на основе графа знаний и его встраиваний (KGEu). Эти встраивания используются в нейросети для работы с изображениями.	Более сложная модель, которая дополнительно использует обратную связь между предсказаниями нейросети и графовыми встраиваниями (KGE_s).



Цель схемы:

- Объединить информацию из **графов знаний** и **данных изображений** для улучшения предсказаний.
- Использование нейросетей (DNN) и встраиваний графов (KGE) позволяет лучше анализировать и обобщать данные.

Датасеты

- Cifar10
- Mini-ImageNet и его производные:
- Mini-ImageNet (Mini)
- ImageNetV2 (V2)
- ImageNet-Sketch (Sketch)
- ImageNet-R (R)
- ImageNet-A (A)



SupSSL	Horizontal_bar	Toucan	Miniskirt	Oboe	Frying_pan	Goose	Catamaran
GAE Visual	Horizontal_bar	Toucan	IPod	Hair_slide	Wok	Goose	Catamaran
GAE Taxonomical	House_finch	American_robin	King_crab	Walker_hound	Frying_pan	House_finch	Yawl
GAE Functional	Goose	Toucan	King_crab	Malamute	Frying_pan	Goose	Catamaran
GAE Generic	American_robin	Toucan	Clog	Dalmatian	Wok	Goose	Catamaran
GAT Visual	House_finch	Toucan	King_crab	Tibetan_mastiff	Wok	Goose	Catamaran
GAT Taxonomical	American_robin	Toucan	King_crab	Malamute	Frying_pan	Goose	Catamaran
GAT Functional	Toucan	Toucan	King_crab	Stage	Frying_pan	Goose	Catamaran
GAT Generic	Horizontal_bar	Toucan	Pencil_box	Toucan	Frying_pan	Goose	Catamaran

Контекстные прогнозы DNNGAE (GAE) и DNNGAT (GAT) и их контекстный вид на Mini-ImageNet. Контекстный вид влияет на представление изображения и, следовательно, на окон чательный прогноз для одного и того же входного изображения.

Спасибо за внимание!