

ПРОЦЕСС КАТЕГОРИЗАЦИИ ФОТОГРАФИЙ: СТРАТЕГИИ, МЕТОДЫ, АРХИТЕКТУРА

Артихович Н.С., студент

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Парамонов А.И. – канд. техн. наук, доцент, зав. каф. ИСиТ

Исследуются методы категоризации фотографий, включая обзор применения различных моделей машинного обучения для этой задачи. Описаны ключевые аспекты для создания системы категоризации, предлагающей пользователю классификацию изображений на основе его потребностей.

Категоризация фотографий представляет собой сложную задачу, которая требует использования интеллектуальных алгоритмов и машинного обучения. Основная проблема заключается в том, что фотографии могут быть различными по содержанию, стилю, качеству и другим параметрам. Более того, это задача, которая включает в себя как различие очевидных визуальных отличий (например, фотография горы и фотография пляжа), так и более тонкие различия, которые могут быть менее очевидны (например, различные виды диких животных). Самообразовательные системы обычно требуют большого количества данных для "обучения", что добавляет дополнительное время и стоимость для процесса. Кроме того, часто бывает сложно точно определить, как система должна классифицировать каждую фотографию, так как многие изображения могут относиться к нескольким категориям одновременно.

Категоризация фотографий — это процесс, при котором изображения классифицируются в различные группы на основе их содержания. Существуют различные методы категоризации фотографий, среди которых можно выделить [1]:

1. Ручная категоризация: классический метод, который включает в себя процесс просмотра фотографии человеком и присвоения ей определенных меток или категорий. Хотя этот метод позволяет достигнуть высокой степени точности, он трудоемок и может быть неэффективен при обработке больших объемов данных.

2. Машинное обучение: с помощью этого метода компьютеры "учатся" классифицировать изображения на основе большого количества информации, которую они обрабатывают. Такие методы иногда совмещаются с глубоким обучением, что позволяет модели обучаться на более сложных и абстрактных признаках.

3. Глубокое обучение и сверточные нейронные сети (CNN): Глубокие нейронные сети, особенно сети свертки, стали очень популярными в области категоризации изображений из-за их способности эффективно обрабатывать и анализировать изображения. CNN может выявлять важные признаки изображения и классифицировать его с помощью этих признаков.

4. Семантическая категоризация: метод, который включает внедрение метаданных в изображение, такие как геолокация, время съемки и теги, что позволяет еще более точно классифицировать изображение.

5. Кластеризация: тип обучения без учителя, который предполагает группировку изображений в категории на основе их визуального сходства, без предварительного определения меток категорий.

Модели, которые уже были обучены на больших наборах данных, таких как ImageNet, считаются предварительно обученными. Они уже распознают тысячи различных объектов и понимают разные особенности изображений. Использование предварительно обученной модели обычно упрощает процесс обучения и улучшает его точность, поскольку нам не нужно начинать обучение с нуля. От нас требуется лишь дообучить модель для наших конкретных задач, научив ее распознавать новые классы на основе уже выученных признаков [2].

Метод быстрой категоризации фотографий или изображений с помощью легковесных моделей представляет собой инновационный подход к обработке и интерпретации информации на изображениях. Благодаря его использованию, процесс распознавания и классификации элементов на фотографии значительно увеличивается в скорости и эффективности. Принцип действия описан на Рисунке 1.

В основе этого метода лежит использование максимально простых и быстрых моделей, таких как "inceptionv3 finetuned parameters over 1M". Суть данной модели заключается в детализированной настройке параметров, позволяющей достичь высокой степени точности распознавания объектов на фото. Это модель глубокого обучения, которая отличается небольшим количеством памяти, необходимой для работы алгоритма, и высокой скоростью обработки информации.

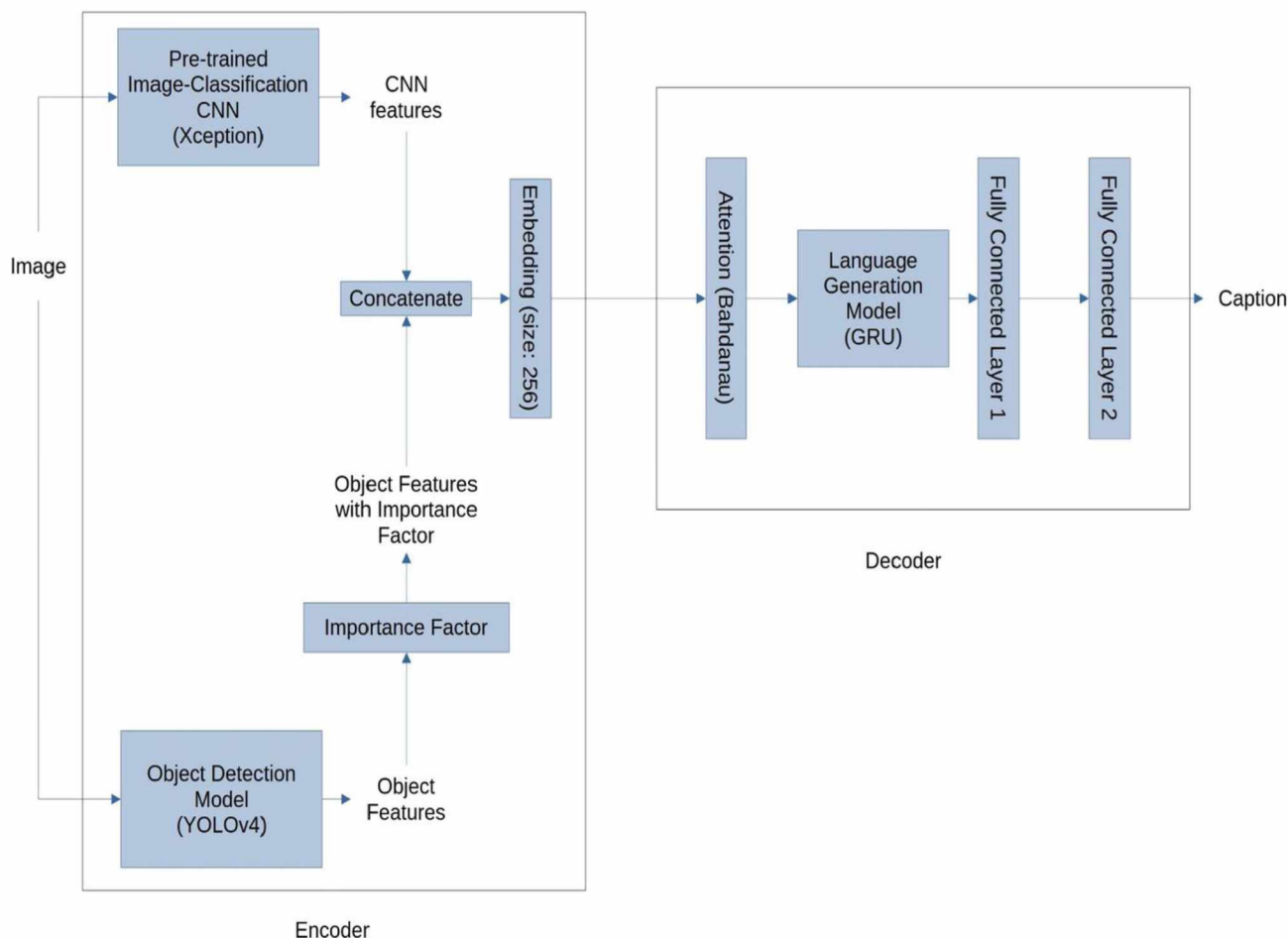


Рисунок 1 – Принцип работы распознавания изображений

Однако, данная модель используется не в одиночку. Существенную роль играет сбор информации об изображении от пользователя. К примеру, место, в котором была сделана фотография, время снимка, наличие людей на фото, альбом, где хранится фотография, и даже текст, присутствующий на изображении — всё это может быть использовано для более точного и полного описания фотографии.

После сбора всей дополнительной информации она посылается в модель Language Learning Model (LLM) с запросом "Сгенерируй подробное описание". Здесь происходит финальный этап - генерация полного и исчерпывающего описания изображения.

Такой подход гармонично объединяет два ключевых аспекта: быстроту и точность распознавания изображений с помощью легковесных моделей, и учет дополнительной информации о фото для более полного описания. В результате пользователь получает моментальное и детализированное описание любого фото или изображения, что расширяет границы возможного использования таких систем, например, для организации фотоархивов, диагностики по медицинским изображениям, систем безопасности, социальных сетей и много другого [3].

Список использованных источников:

- [1] Deep Learning Approaches on Image Captioning: A Review – Режим доступа: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00571-w>.
- [2] Image captioning model using attention and object features to mimic human image understanding. – Режим доступа: <https://arxiv.org/pdf/2201.12944.pdf>.
- [3] Automatic image captioning combining natural language processing and deep neural networks. – Режим доступа: <https://www.sciencedirect.com/science/article/pii/S2590123023002347>.