CNN Fixations

"An unraveling approach to visualize the discriminative image regions."

discriminative img regions – special regions which provides special information abt the obj.

Convolutional Neural Networks

2012 – AlexNet, 8 layers, 60 m parameters

↓

2015 – ResNet, 100s of layers, 1,7m parameters

Disadvantage: CNN is a black box.

data → [?] → result

To understand → look at the important img regions that influence their predictions

 → bird

① Вступление

1) Основная идея:

Авторы предлагают способ визуализации того, как CNN принимает решения. Они используют информацию о взаимосвязях м/у выходами нейронов на соседних слоях сети. Когда CNN делает предсказание, активируются опр. нейроны. Это позволяет понять, какие активации на предыдущих слоях сети привели к активации на текущем слое. Такие действия происходят на каждом уровне сети, от softmax (вых. слой) до входного изображения.

Процесс работы:

- выбирается нейрон на каком-либо слое
- Система находит активации на предыдущих слоях, которые поддерживают активацию выбранного нейрона
- повторяется процесс до момента, когда дойдём до вх. изображения
- Т.о., метод помогает определить конкретные участки изображения, которые явл. ответственными за предсказание модели.

Применение:

- метод наглядно даёт понять, например, какие участки изображения позволяют понять, на изображении кошка или собака.
- генерация подписей для изображения.

Преимущества метода:

- модель делается более понятной и прозрачной. Визуализация помогает понять, почему сеть приняла такое решение.
- метод можно использовать не только для финальных активаций, но и для любых др. нейронов на др. уровнях.
- точная покадровая локализация объектов на изображении.

② Сопоставляющая работа.

- Бо́льшая часть работ – градиентный подход: находят области изображения, которые могут улучшить прогнозируемую оценку для выбранной категории.
- Ход и др: карты активации могут быть получены путём объединения карт признаков перед слоем GAP (Global Average Pooling) в соответствии с весами, соединяющими слой GAP с активацией класса в слое классификации.
- Подход, основанный на оценке того, как изм. прогноз, если ф-ция отсутствует.
- → • В отличие от др. работ, подход в статье находит "ответственные" местоположения пикселей, просто распространяя базовые операции прямого прохода через сеть.

Пояснения по работе метода:

Работа начинается с нейрона, который отвечает за предсказанную категорию (например, "кошка"). Далее определяются нейроны, которые активировались на предыдущих слоях. Этот процесс называется "разворачивание" активации (unraveling).

Процесс вывода результата:

Вместо того, чтобы восстанавливать активации (как это делают др. методы), их метод выдаёт бинарный результат на каждом слое сети => становится понятно, какие нейроны были нужны, а какие нет. После этого создаётся тепловая карта (heat map) при помощи размытия бинарного результата Гауссовским размытием (мат. фильтр, который сглаживает картинку).

Простота метода:

Метод не требует настройки параметров (гиперпарам-ов) или др. сложных эвристик (правил).

① Авторы: L.K. Hansen, E.A. Hendricks, N.A. Lydersen, A. Blanchard

Учреждение: Technical University of Denmark, Image and Signal Processing Group

② Год: 2016 year

③ Название сети: CNN (Convolutional Neural Network) (что)

④ Назначение сети: (зачем-задача, входные и выходные данные) Сеть CNN используется для классификации изображений и распознования объектов. Цель состоит в том, чтобы визуализировать дискриминационные области изображения (фиксации), которые CNN использует для принятия решений.

Входные данные: набор изображений, которые используются для обучения и тестирования CNN.

Изображения проходят через несколько слоёв сети (свёрточные слои, пулинг и полносвязные слои), где на каждом уровне происходит обработка и извлечение признаков.

Выходные данные: предсказания сети, то есть метки классов изображений, а также визуальные карты фиксаций, показывающие, какие области изображения были наиболее значимы для принятия решений сети.

Дополнительно с помощью предложенного метода "fixation mapping" выводятся карты фиксаций.

(5) Мотивация авторов к разработке (почему): Мотивация заключается в том, чтобы лучше понять внутр. механизмы работы CNN, которые являются "чёрными ящиками". Несмотря на высокую точность CNN при распознавании объектов и классификации изображений, важно раскрыть, какие части изобр-ий опр. выводы сети.

Это помогает:

- улучшить доверие к работе НС, особенно в ответственных приложениях, таких как мед. и беспилотные системы.
- опр. слабые места сети, которые можно улучшить для более точной классификации.
- ускорение разработки новых арх. CNN

⑥ Состав и отличие от базовой архитектуры (как решают):

В осн. ленит стандартная структура CNN, но авторы добавили метод визуализации "fixation mapping". Этот метод позволяет на каждом слое сети выделить те области изображения, которые оказывают наиб. влияние на предсказания сети.

*Важная особенность: не требует изм. в базовой архитектуре CNN, но существенно расширяет понимание того, как сеть "видит" из-я на разных уровнях. В отличие, например, от метода Grad-CAM, фиксации предлагают более прямую визуализацию ключевых областей.

- Grad-CAM – Gradient-weighted Class Activation Method.
- фиксации – показывает ключевые обл. изображения.
- активации – визуализирует отклик нейронов на изображение.

⑦ Качественные и количественные показатели: (как оценивают или сравнивают)

1) Качественные показатели: карты фиксаций
2) Количественные —"—:
- точность классификации CNN, когда сети даются только зафиксированные участки из-я.
- происходит сравнение с исх. изображением и др. методами визуализации.

⑧ Есть ли потенциал развития (плюсы/минусы):

⊕: 1) Возможность улучшения интерпретируемости НС, что важно для таких приложений с использованием НС, например для классификаций изображений или в мед. отрасли для анализа снимков.
2) Простота интеграции с сущ. арх-ми CNN.
3) Возможность исп-я метода для улучшения обучения НС, направляя внимание на ключ. эл-ты изображений.

⊖: 1) Сложно обобщить для более сложных и разнообр. наборов данных
2) Ограничение на типы сетей. Метод лучше всего подходит для CNN.