The Emergence of Convolution

Pre-work: Computer Vision

Agenda

- Feature Extraction

- Conventional Feature Extraction Techniques

- Convolution over SIFT and HOG

Feature Extraction

Feature Extraction in Images

- **What is a feature** (in an image)?
  - Features are various forms of information that can be gained from an image.
    For example: Fig 1 has various features, such as shape, size, color, edges, and background. These features play a crucial role in helping perform several tasks in computer vision like image classification, object detection, scene detection etc.

- The features required by each task in computer vision are completely task-dependent, and each task might not require all the features available in an image.
  For example: We can still identify that Fig 2 is an image of a duck (even though it is harder to do so than for Fig 1) without necessarily knowing the color or background of the image, but only from looking at the edges, the presence of water and the shape of the bird inside the image.

Fig 1

Fig 2

Feature Extraction in Images

- Hence, we need methods to extract **just the essential information** contained in images and ignore the rest depending on different tasks. This step is called **feature extraction**.

- The features we obtain from feature extraction are crucial in enhancing the model's   performance.

- Feature extraction methods help in converting images to certain feature vectors   of   fixed   sizes.

- Some of the most important feature extraction techniques to have been used with images are:

    - HOG (Histogram of Oriented Gradients)
    - SIFT (Scale Invariant Feature Transform)
    - Convolution   Operations

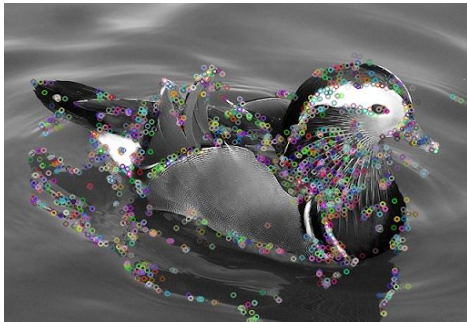# Conventional Feature Extraction Techniques

Conventional Feature Extraction Techniques

The two most historically-important, manual feature extraction techniques have been:

1. **HOG** (Histogram of Oriented Gradients)
2. **SIFT** (Scale Invariant Feature Transform)

HOG	SIFT
HOG mainly focuses on the structure and shape of an object.	In SIFT, image content is converted into local feature coordinates that are not affected by rotation, scaling, or other image manipulations.
HOG is different from only detecting edges, as it also identifies the magnitude and direction of edges in the image.	SIFT assists in locating the local features of an image, often known as image keypoints.

Conventional Feature Extraction Techniques

HOG	SIFT	Original image
HOG calculates the magnitude and direction of edges in each region. The Orientation is the direction and the Gradient is the magnitude of the pixel values of the image.	The key points obtained from SIFT can be utilised for picture matching, object detection, scene detection, and other computer vision applications.	

Disadvantages of Conventional Feature Extraction Techniques

- **Why was SIFT preferred over HOG?**
  - SIFT features, as opposed to HOG features, have the advantage of being unaffected by the image's size or orientation.
  - SIFT has a better accuracy than HOG for detecting features in an image.
  - HOG is not scale and rotation invariant, whereas SIFT shows those properties.
- However, **both SIFT and HOG showed certain disadvantages** in the efforts to apply them as general feature extraction techniques for all images:
  - Both SIFT and HOG are quite slow and computationally expensive.
  - They are also somewhat mathematically complex in their working.
  - In addition, HOG does not work well with lighting changes and blurring in the images.

Convolution over SIFT and HOG

Convolution

- Convolution is a **specialized linear operation** on an image, that represents an efficient way of extracting image features and reducing the dimensions of an image.
- Convolution consists of a set of filters called convolution layers, that perform convolution operations on images.
- We use multiple filters to perform convolution operations on an image, and try to extract various kinds of features (pertaining to each kind of filter) from a single image.
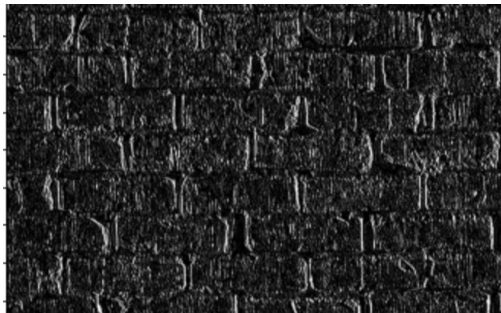
    For example: Let's assume that we need to use convolutions to extract features from this image of a brick wall.
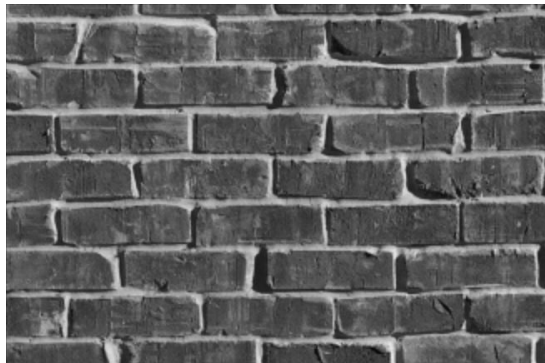
Convolution

- We may be interested in each of the following feature extraction tasks:
  - to extract **all the vertical edges** from the image
  - to extract **all the horizontal edges** from the image
  - to **blur/sharpen** the image
  - To **highlight/focus on certain places** of image

- As seen from the below examples, convolution filters can help us achieve any of the feature extraction tasks we may require for our images.

Vertical edge detection   Horizontal edge detection

Convolution over SIFT and HOG

Why is Convolution better than SIFT and HOG?

- Convolutional feature detectors are **highly trainable and adaptable**, allowing them to achieve higher accuracy levels in comparison to SIFT and HOG for the task at hand.

- **Convolutions excel at learning low-level features** of an image in a much better way than SIFT and HOG, and they do so **without the overhead of the hand-coded feature engineering** which is usually required for SIFT and HOG.

- Apart from learning low-level features, hierarchical combinations of convolutions are quite effective in learning important **high-level features** as well.

  Example: For images of human faces, convolutional layers would easily learn to understand more complex shapes such as the eyes, the ears, the nose or the mouth.

Convolution over SIFT and HOG

Why is Convolution better than SIFT and HOG?

- In 2012, **AlexNet**, a Convolutional Neural Network (CNN) architecture, based fundamentally on the principle of convolutional filters, handily won the famous ImageNet competition - **outperforming the runner-up by over 10 percentage points.** Although convolutions were already known in literature from the work of Yann LeCun, this breakthrough is what drew attention from the whole technology industry to the power of convolutions in image feature extraction.

- It soon became clear to machine learning practitioners that hierarchical combinations of convolution filters achieved **superior and far more generalizable results** in image feature extraction than SIFT and HOG. This is the fundamental driver behind the emergence of convolutions as part of Convolutional Neural Networks (CNNs), which have become a staple in state-of-the-art deep learning models for computer vision over the last decade.

Thank You