**Convolutional Neural Network (CNN)**

1. 개요
2. 배경

일반적으로 DNN(deep neural network, especially for fully-connected neural network)은 1차원 형태의 데이터를 사용한다. 때문에 (예를들면 1028x1028같은 2차원 형태의)이미지가 입력값이 되는 경우, 이것을 flatten시켜서 한줄 데이터로 만들어야 하는데 이 과정에서 이미지의 공간적/지역적 정보(spatial(topological)/local information)가 손실되게 됩니다. 또한 추상화과정 없이 바로 연산과정으로 넘어가 버리기 때문에 학습시간과 능률의 효율성(from network size, number of parameters)이 저하됩니다 [1][2].

1. 개념

이미지에서의 특정 위치에 있는 픽셀들은 그 주변에 있는 일부 픽셀과 correlation 높고, 거리가 멀어지면 멀어질수록 그 영향은 감소함.

(-> 공간적/지역적 정보 손실 감소)

이미지 전체 영역에 대해 서로 동일한 연관성(중요도)로 처리하는 대신에 특정 범위에 한정에 처리한다면 훨씬 효과적일 것이 아이디어가 제시됨; sparse interactions(sparse interactions은 원본 이미지보다 크기가 작은 필터를 사용함으로써 실현됨)

(-> 학습시간, 네트워크 크기, 파라미터 수 감소)

CNN은 이미지를 날것(raw input) 그대로 받음으로써 공간적/지역적 정보를 유지한 채 특성(feature)들의 계층을 빌드업합니다.

(-> 공간적/지역적 정보 손실 감소?)

CNN의 중요 포인트는 이미지 전체보다는 부분을 보는 것, 그리고 이미지의 한 픽셀과 주변 픽셀들의 연관성을 살리는 것입니다.

1. 특징

Locality(Local Connectivity)

* CNN은 Local 정보를 활용하여 인접한 신호들에 대한 correlation 관계를 비선형 필터를 적용하여 추출 함. • 이런 필터를 여러 개 적용하면 다양한 local 특징을 추출할 수 있으며 • 이러한 과정을 거치면서 영상의 크기는 줄어들며 global feature를 얻을 수 있게 됨

Shared weights

* Filter를 반복적으로 적용함으로써 변수를 획기적으로 줄일 수 있음 • Topology 변화에 무관한 항상성(invariance)를 얻을 수 있음

1. 구조

텍스트, 도표, 스크린샷, 평행이(가) 표시된 사진

자동 생성된 설명

1. CNN은 이미지의 특징을 추출하는 영역과 이미지의 클래스를 분류하는 영역으로 구성됨
2. 이미지의 특징을 추출하는 영역(feature learning: feature extraction)은 convolution layer와 pooling layer를 다층으로 구성함; convolution layer는 입력 데이터에 필터를 적용한 후 활성화 함수를 반영하는 필수요소임. Pooling layer는 선택적인 요소임.
3. 이미지의 클래스를 분류하는 영역(classification)은 fully connected layer로 구성함. fully connected layer는 cnn의 마지막 부분에 위치함
4. Flatten layer는 이미지 형태의 데이터를 배열형태로 만드는 레이어임; 이미지의 특징을 추출하는 영역과 이미지의 클래스를 분류하는 영역 사이에 위치함
5. Convolution layer [3]

* Convolution layer는 convolution 연산을 통해서 이미지의 특징을 추출하는 역할을 수행한다
* 컨볼루션(합성곱) 연산
* 입력 데이터에 필터(커널)를 적용한다.
* 입력, 필터, 출력

가로 세로 방향의 shape

필터: 이미지의 특징을 찾아내기 위한 공용 파라미터

표기: (height, width)

* 입력(이미지): (p,q) 행렬, 필터: (n,m) 행렬, 출력(이미지): (s,r) 행렬
* nxm 행렬 크기의 필터로 pxq 행렬 크기의 이미지를 처음부터 끝까지 겹치며 훑으면서 nxm 크기의 겹쳐지는 부분의 각 이미지와 커널의 원소의 값을 곱해서 모두 더한 값을 출력으로 연산하는 것이다. 이를 통해 이미지의 feature map을 만들 수 있다

텍스트, 도표, 평면도, 라인이(가) 표시된 사진

자동 생성된 설명

도표, 평면도, 라인, 기술 도면이(가) 표시된 사진

자동 생성된 설명

1. Pooling layer [4]

* 이미지 크기를 줄이고 특정 feature를 강조하는 역할
* 이미지 크기가 줄어들기 때문에 인공신경망의 모델 파라미터 또한 감소한다 따라서 Pooling layer를 적용함으로써 CNN 학습시간을 감소시킬수 있고 오버피팅 문제도 완화할 수 있음
* 이미지 데이터의 특징은 인접 픽셀들간의 유사도가 매우 높다. 따라서 이미지는 픽셀 수준이 아니라, 특정 속성을 갖는 선택 영역 수준으로 표현될 수 있다는 특징을 바탕으로 설계됨
* 선택영역 내부에서는 픽셀들이 이동 및 회전 등에 의해 위치가 변경되더라도 출력값은 동일하다. 따라서, Pooling을 적용함으로써 이미지를 구성하는 요소들의 이동 및 회전 등에 의해 CNN의 출력값이 영향을 받는 문제를 완화할 수 있음
* 처리방법
* Max Pooling
* Average Pooling
* Min Pooling
* Max Pooling

선택영역에서 가장 큰 값을 해당영역의 대표값으로 설정

직사각형, 도표이(가) 표시된 사진

자동 생성된 설명

1. Fully connected layer

* Flatten layer + softmax layer
* Flatten layer: 이미지 데이터를 flatten(배열형태)하게 만듬으로써 이미지 데이터의 shape만 변경함; 이미지의 특징추출 부분과 분류부분 사이에 위치함. fully connected layer의 입력
* Softmax layer: 이미지에 대한 activation을 수행함으로써 classification을 수행함

1. 주요 용어 및 개념
2. **Spatial hierarchies and patterns in image data** refer to the organization and arrangement of visual information in different levels of abstraction within an image. spatial hierarchies and patterns in image data describe how a CNN learns to understand the spatial relationships and structures within an image at different levels of abstraction. This hierarchical learning process allows the network to gradually recognize and combine local features into more complex patterns, enabling it to perform tasks such as image recognition, object detection, and image classification.

* **Spatial Hierarchies:**
  + "Spatial" refers to the space or layout of the image, involving the position and arrangement of pixels.
  + "Hierarchies" indicate a structure with different levels of abstraction or detail.
  + In the context of image data and Convolutional Neural Networks (CNNs), spatial hierarchies mean that the network learns to recognize features at different scales and complexities. Lower layers in a CNN tend to capture basic and local features, such as edges, textures, and simple shapes. As you move up through the layers, the network combines these basic features to recognize more complex and abstract patterns that represent larger and more meaningful structures in the image.
* **Patterns in Image Data**
  + "Patterns" refer to recognizable arrangements or combinations of features in the image.
  + In image data, patterns can be anything from simple textures or shapes to complex structures like objects or scenes. CNNs are designed to automatically learn and extract these patterns from the input images during the training process. The convolutional layers in a CNN are particularly effective at capturing patterns by using filters or kernels that slide across the image, identifying local features and assembling them into higher-level representations.

1. **Tensor**

a mathematical object representing a multi-dimensional array of numerical values. Tensors can be scalars (single numbers), vectors (1D arrays), matrices (2D arrays), or higher-dimensional arrays.

Type of tensors

**Scalar:** A single numerical value.

**Vector:** An ordered array of numbers (1D tensor).

**Matrix:** A 2D array of numbers.

**Higher-dimensional tensors:** Tensors with more than two dimensions. For example, a 3D tensor can be thought of as a cube of numbers, and a 4D tensor can be visualized as a cube of cubes.

In machine learning, tensors are crucial for representing and manipulating data because they allow for the efficient parallelization of operations, which is essential for training large neural networks.

1. **Channel**

In computer vision, a "channel" typically refers to a specific component of an image or a feature map. The concept of channels is often associated with color images and convolutional neural networks (CNNs). Let's explore both contexts:

1. **Color Images:**
   * In the context of color images, the term "channel" refers to one of the color channels that compose the image. The most common color representation is the RGB (Red, Green, Blue) model, where an image is composed of three channels: one for red, one for green, and one for blue. Each channel represents the intensity of its respective color.
   * For example, a color image with shape (height, width, 3) has three channels, and each channel is a 2D array representing the intensity of a specific color at each pixel.
2. **Convolutional Neural Networks (CNNs):**
   * In the context of CNNs, a channel is a set of feature maps produced by applying a convolution operation to the input image or a previous layer's feature maps. Each channel in a layer captures specific patterns or features.
   * In the first layer of a CNN, each channel may represent simple features like edges or textures. As you go deeper into the network, channels typically capture more complex and abstract features.
   * For example, in a convolutional layer, the input may have multiple channels (if it's not the first layer), and the layer applies multiple filters to produce an equal number of output channels. Each output channel represents the response of a specific filter to the input.

In summary, in computer vision, a channel can refer to the color channels in a multi-channel color image or the feature channels in the context of convolutional neural networks. Understanding and manipulating channels are crucial for various image processing and computer vision tasks.

1. **Fully connected layer (전결합계층) vs Convolution layer (합성곱계층) [5]**

**텍스트, 도표, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명**

**텍스트, 도표, 라인, 폰트이(가) 표시된 사진

자동 생성된 설명**

**텍스트, 도표, 라인, 평면도이(가) 표시된 사진

자동 생성된 설명**

**(3) Conv 연산 예시 [5]**

**텍스트, 도표, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명**

**(4) FC in CNN 연산 예시 [4]**

**텍스트, 도표, 라인이(가) 표시된 사진

자동 생성된 설명**

**텍스트, 도표, 친필, 폰트이(가) 표시된 사진

자동 생성된 설명**

1. **Feature map [5]**

**텍스트, 스크린샷, 도표, 폰트이(가) 표시된 사진

자동 생성된 설명**

1. **Stride [5]**

**라인, 사각형, 직사각형, 도표이(가) 표시된 사진

자동 생성된 설명**

1. **Padding [5]**

**텍스트, 스크린샷, 번호, 도표이(가) 표시된 사진

자동 생성된 설명**

1. **Pooling [5]**

**직사각형, 도표이(가) 표시된 사진

자동 생성된 설명**

**텍스트, 스케치, 도표, 스크린샷이(가) 표시된 사진

자동 생성된 설명**

**텍스트, 스크린샷, 직사각형이(가) 표시된 사진

자동 생성된 설명**

**텍스트, 스크린샷, 도표이(가) 표시된 사진

자동 생성된 설명**

1. **Flatten[5]**

텍스트, 도표, 스크린샷, 디자인이(가) 표시된 사진

자동 생성된 설명

1. 예시

* LeNet-5, AlexNet, VGG-16, GoogLeNet(Inception)
* ResNet, DenseNet

1. 참고문헌

[1] <https://mijeongban.medium.com/%EB%94%A5%EB%9F%AC%EB%8B%9D-%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-cnn-convolutional-neural-networks-%EC%89%BD%EA%B2%8C-%EC%9D%B4%ED%95%B4%ED%95%98%EA%B8%B0-836869f88375>

[2] <http://infosec.pusan.ac.kr/wp-content/uploads/2017/11/CNN-and-RNN-%EC%9D%B4%EB%A1%A0.pdf>

[3] <https://wikidocs.net/64066>

[4] <https://rubber-tree.tistory.com/entry/%EB%94%A5%EB%9F%AC%EB%8B%9D-%EB%AA%A8%EB%8D%B8-CNN-Convolutional-Neural-Network-%EC%84%A4%EB%AA%85>

[5] https://seunghan96.github.io/dl/cv/ppt/(%EB%B0%9C%ED%91%9C%EC%9E%90%EB%A3%8C)CNN/