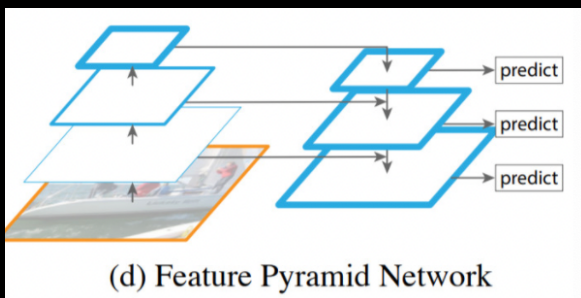


< 1. 서론 >

- VGGNet은 심층 컨볼루션 신경망의 개념으로 설계되어 대규모 이미지 분류에 좋은 성능을 가지고 있지만, 위치 인식에 약간의 오류가 있다.
- 위치 인식 오류를 줄이고 높은 수준의 특징맵을 추출하기 위한 독특한 네트워크 구조인 PPCNN(피라미드 컨볼루션 신경망)을 활용.
- 작은 물체를 정확하게 감지하기 위해서는
 - 1. 물체와 배경의 정확한 구분이 필요하다
 - 2. 여러 물체들을 다르게 구분하기 위해 bounding box를 정확하게 예측해야한다.
- 문제는 대부분의 detector는 가까이 있고 positive한 물체를 우선적으로 인식한다.
- PPCNN (Pyramid Pooling Convolutional Neural Network)를 통해 layer의 복잡도를 없애고, 높은 품질의 feature map을 추출해낸다.
- PPCNN은 VGGNET에 U자 형태의 feature pyramid network “FPN” (보조 역할)가 첨가된 형태이다.
- Feature pyramid network는 높은 품질의 feature map을 추출하는 용도로 활용된다.
-

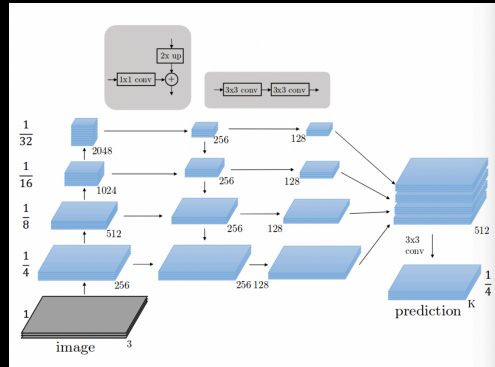
< 2. Related Work >

< 2.1 Over View of FPN >



- 다양한 크기의 객체를 인식하는 것은 object detection의 핵심 문제. 그러나 기존 모델들은 추론 속도가 너무 느리고, 메모리를 지나치게 많이 사용하는 문제점이 있다.
- FPN은 컴퓨팅 메모리를 적게 차지하고, 다양한 크기의 객체를 인식하는 방법을 제시한다.
- 기본적으로 convolutional network에서 입력층에 가까울 수록 feature map은 높은 해상도를 가지며, 멀어질 수록 (가장자리, 곡선)은 낮은 특징과 낮은 해상도를 갖는다.

< 2.2 Structure of FPN >

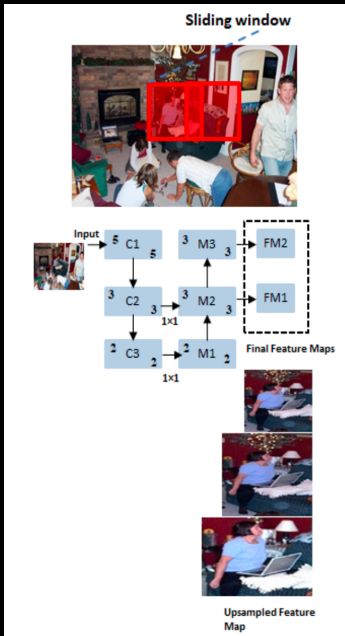


- 임의의 크기의 single scale 이미지를 convolutional network에 입력하여, 다양한 크기의 feature map을 추출 및 수정하는 네트워크.
- 계층적인 convolutional pyramid layer
- Pyramid : 이는 convolutional network에서 얻을 수 있는 서로 다른 해상도의 feature map을 쌓아 올린 형태이다 또한 level이란 피라미드 각층의 feature map을 의미한다. 여러 level이 쌓아올려진게 pyramid.
- Convolutional layer는 여러 convolution filter들로 쌓여있는 형태이다. 이때 각 convolution filter 들을 low level부터 high level 이미지들의 feature를 추출한다.
- Pyramid 구축을 하는 과정에서 bottom up pathway, top-down pathway, lateral connection 과정을 거친다.
- Bottom-up pathway는 이미지를 convolutional network에 입력하여, forward pass하면서 2배씩 작아지는 feature map을 추출하는 과정이다. 이때 각 stage의 마지막 layer이 가장 깊음으로 feature map이 가장 좋은 정보를 가지게 되어, 마지막 Layer의 feature map을 추출하게 된다.
- Top-down pathway는 각 pyramid level에 있는 feature map을 2배로 upsampling 하며 channel 수를 맞춰주는 역할을 한다. 각 pyramid level의 feature map이 2배로 upsampling을 하며, 아래 pyramid level의 feature map과 동일한 크기를 가지게 됩니다. (Nearest neighbor up sampling 활용).
- Lateral Connection :upsampling된 feature map과 아래 level의 feature map의 element wise addition 연산을 수행.
- 해당 과정들을 통해 각기 다른 scale에서의 feature map들을 추출할 수 있다.

< 3 Proposed Algorithm >

< 3.1 PPCNN >

- pyramid pooling network를 통해 출력을 위한 다양한 크기의 feature들을 추출하고, 이를 통해 같은 크기의 feature map들을 얻어낼 수 있다.



- PPCNN의 구조. (VGGNET을 활용)
- 입력 feature 들을 3가지 크기의 형태로 정해준다 {C1,C2,C3 } = { 32,64,128}
- Feature scaling은 stride = 2, stride = 4의 max pooling으로 진행된다.
- Mid layer에서 추출된 입력 feature들은 pyramid pooling network (upsampling connection을 포함한 merging cell을 가짐). 로 보내진다. 이후 pyramid network는 증강된 {M1,M2,M3} multi scale의 feature를 내 보낸다.
- Pyramid pooling network이 동일한 규모의 feature layer로 구성되는 점을 고려하면, FPN은 더 나은 효율성을 위해 반복적으로 누적될 수 있다.

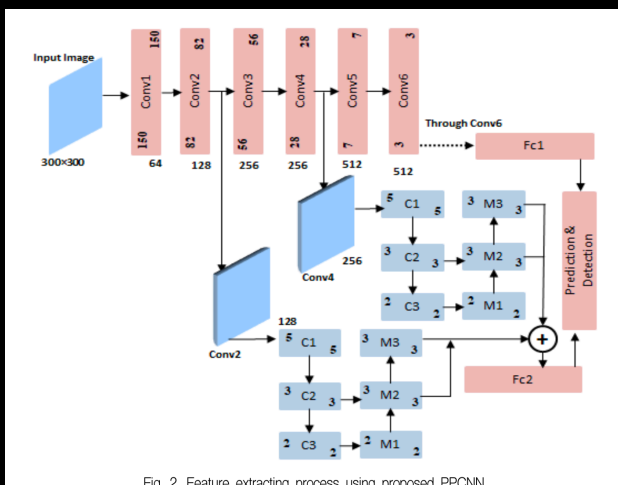


Fig. 2. Feature extracting process using proposed PPCNN

- VGGNet은 localization 과 feature extracting 문제가 있고 이를 해결하기 위해 pyramid 구조를 활용하는 것이다.
- 첫 두 layer는 feature 보존을 하는 역할을 수행한다
- 세번째 layer는 이전 two fully connected layer과 비교를 통해 객체를 분류하는 1000개의 channel을 가졌다.

- 좌측의 그림에 보이듯이 위 구조는 FPN의 bottom-up, top down pathway를 가지고 있다. High quality feature map들을 추출하여 FC2에 보존한다.

$$P_n^{ij} = \frac{1}{(H_n \times W_n)} \cdot \sum_i^{H_n} \sum_j^{W_n} f_n(i, j) \quad (1)$$

해당 공식은 average pooling feature data of PPCNN network를 나타낸다.

- 참고자료 :
- <https://herbwood.tistory.com/18> (FPN)