김일성종합대학학보
정보과학
주체110(2021)년
제67권 제1호

중첩신경망을 리용한 통합구조형 대상검출방법

홍 광 철

경애하는 **김정은**동지께서는 다음과 같이 말씀하시였다.

《첨단과학기술분야에서 세계적경쟁력을 가진 기술들을 개발하기 위한 투쟁을 힘있게 벌려야 합니다.》

선행연구[1, 2]에서는 입력되는 화상에서 대상으로 판단될수 있는 부분령역들을 추출하는 대상후보령역추출단계와 부분화상들을 분류기를 통과시켜 대상인가 아닌가를 분류하는 방법을 제안하였다.

론문에서는 입력되는 화상을 하나의 중첩신경망만을 통과시켜 검출하려는 대상위치와 크기, 대상의 클라스확률을 얻어내는 실시간적인 대상검출방법을 제안하였다.

1. 대상검출과정

대상검출문제를 공간적으로 분산되여있는 대상의 경계4각형과 대상클라스확률을 하나의 신경망으로 얻어내는 회귀문제로 보고 옹근화상(입력화상을 그대로 리용한다는 의미)에서 한번의 평가로 위치와 확률을 예측한다.

검출과정을 그림 1에 보여주었다.

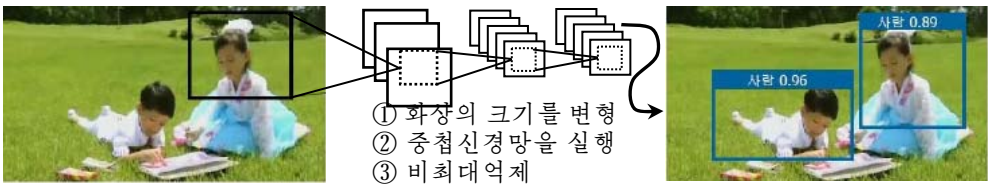

그림 1. 검출과정

그림 1에서는 제안한 대상검출방법의 전반적인 과정을 보여주었다. 1개의 신경망으로 여러개의 대상령역을 예측하고 동시에 그것들의 클라스확률도를 예측한다.

입력화상을 $S \times S$ 살창으로 나눈다. 만일 대상의 중심이 어느 한 살창에 떨어지면 그 살창은 그 대상을 검출하였다는것을 의미한다.

매개 살창은 B개의 대상과 매 대상의 믿음도를 예측한다. 이 믿음도는 살창이 얼마의 정확도로 대상을 예측하고있는가를 반영한다.

믿음도는 $\Pr(\text{Object}) \times \text{IOU}_{\text{pred}}^{\text{truth}}$ 로 정의한다.

만일 살창내에 아무런 대상도 포함하지 않는다면 믿음도는 0일것이다. 믿음도는 예측령역과 정답령역간의 IOU(두 령역의 합과 사귐사이의 비)와 같게 된다.

매 령역은 5개의 값(x, y, w, h, c)을 예측한다. 여기서 (x, y)자리표는 매 살창에서 대상중심의 상태위치를, w, h는 전체 화상에 대한 상대크기를 나타낸다. C는 믿음도로서 예측령역과 정답령역과의 IOU를 나타낸다. 매 살창은 역시 C개의 클라스조건부확률 $\Pr(\text{Class}_i \mid \text{Object})$ 를 예측한다.

이 확률들은 대상을 포함하고있는 살창에서만 필요조건으로 된다. 여기서 예측대상 개수 B는 무시하고 오직 살창에 대한 클라스확률만을 예측한다.

시험때에는 조건부확률들과 개별적인 대상믿음도예측을 곱한다. 이것은 매 대상령역에서 클라스별믿음도를 얻게 한다.

이 값들은 대상령역안에서의 클라스출현확률과 함께 예측령역이 대상일 확률을 동시에 반영한다.

$$\Pr(\text{Class}_i \mid \text{Object}) \times \Pr(\text{Object}) \times \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) \times \text{IOU}_{\text{pred}}^{\text{truth}} \tag{1}$$

체계에서는 검출을 회귀문제로 진행한다. 화상을 작은 살창으로 나누고 매 살창에서 대상의 령역과 령역에서의 믿음도, 클라스확률을 예측한다.

2. 신경망의 설계와 손실함수

신경망의 초기단계의 중첩층들은 화상에서 특징들을 추출하고 전결합층들은 출력확률과 자리표를 예측한다.

신경망의 구조를 그림 2에 보여주었다.

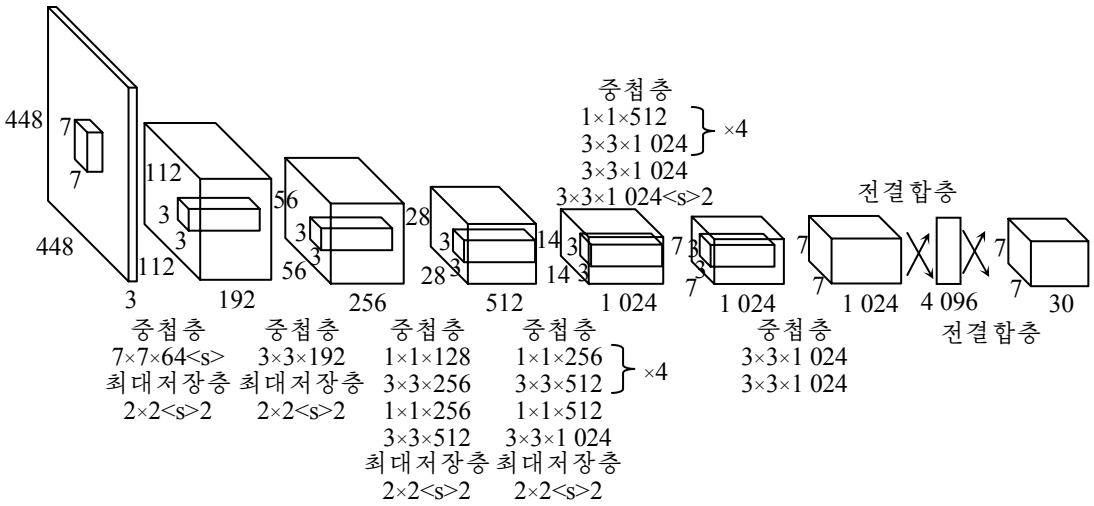

그림 2. 신경망의 구조

대상검출신경망은 화상분류를 위한 신경망중에서 GoogleNet[1]에 기초하여 설계하였다.

제안된 신경망은 24개의 중첩층과 2개의 전결합층을 가지고있다. 마지막층에서는 클라스확률과 위치자리표를 다같이 예측한다. 대상령역의 너비, 높이를 화상의 너비, 높이에 관하여 정규화하여 [0, 1]값구역에 놓이도록 한다. 대상령역의 x, y자리표를 특정한 살창내에서 위치의 편위로 파라메터화하여 x, y값은 [0, 1]구역에 놓이도록 한다.

마지막층에서는 선형활성함수를 리용하고 다른 모든 층들에서는 대상검출분야에서 많이 리용하고있는 개선된 정규선형함수를 리용한다.

$$\phi(x) = \begin{cases} x, & x > 0 \\ 0.1x, & x \le 0 \end{cases} \tag{2}$$

신경망의 학습에서 다음의 다항손실함수를 리용한다.

$$\lambda_{\text{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\text{II}_{ij}^{\text{obj}}(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2+$$
$$+\lambda_{\text{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\text{II}_{ij}^{\text{obj}}\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2+$$
$$+\sum_{i=0}^{S^2}\sum_{j=0}^{B}\text{II}_{ij}^{\text{obj}}(C_i-\hat{C}_i)^2+\lambda_{\text{noobj}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\text{II}_{ij}^{\text{noobj}}(C_i-\hat{C}_i)^2+$$
$$+\sum_{i=0}^{S^2}\text{II}_{ij}^{\text{obj}}\sum_{c\in\text{classes}}(p_i(C)-\hat{p}_i(C))^2$$

손실함수에서 $\text{II}_{ij}^{\text{obj}}$는 i번째 살창에 대상이 있는가 없는가를 나타낸다. $\text{II}_{ij}^{\text{obj}}, \text{II}_{ij}^{\text{noobj}}$는 i번째 살창의 j번째 대상령역을 예측하였는가 못하였는가를 나타낸다.

손실함수에서 첫번째 항은 위치오차를 나타내며 두번째 항은 너비, 높이오차를 나타낸다. 세번째 항과 네번째 항은 대상의 믿음도오차를 나타내고 다섯번째 항은 살창의 클라스확률오차를 나타낸다.

대부분의 살창에는 대상이 포함되여있지 않기때문에 살창들의 믿음도는 대체로 0으로 설정되며 이것은 대상을 포함한 살창들의 믿음도의 그라디엔트를 약화시키는 결과를 준다. 이로부터 위치손실을 증폭시키고 대상을 포함하지 않은 살창들에 대한 믿음도오차를 약화시키기 위해 손실함수의 결수들을 $\lambda_{\text{coord}}=5, \lambda_{\text{noobj}}=0.5$로 설정한다.

신경망의 출력과 대상검출결과를 그림 2에 보여주었다.

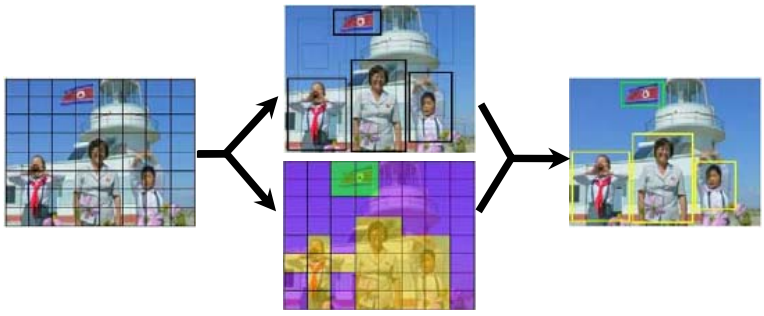

그림 3. 신경망의 출력과 대상검출결과

론문에서 제안한 방법으로 학습된 망모형을 대상검출에 리용할 때 신경망자체가 화상에서 대상의 위치와 크기, 클라스확률을 예측하므로 먼저 입력화상을 정규화하여 신경망평가를 진행한다. 신경망은 구조적특성으로 하여 1개의 화상에서 최대로 98개의 대상을 예측할수 있다. 예측된 대상들가운데서 턱값을 적용하여 믿음도가 높은 대상만을 정확한 대상출력으로 리용한다.

3. 성 능 평 가

대상검출에서의 많은 연구들은 대상후보령역추출속도와 대상분류속도를 높이는데 집중하였다. 그러나 DPM[1]에서는 실지로 초당 30Frame의 속도로 실시간적으로 실행되는

검출체계를 제안하였다. 론문에서 제안한 방법을 2개의 실시간방법(GPU판본에서 30Hz, 100Hz)들과 1개의 비실시간방법(Fast R-CNN[2])과 비교하였다.

Pascal자료모임에서의 성능평가를 표에 보여주었다.

표. Pascal자료모임에서의 성능평가

검출기	mAP/%	FPS/Frame
DPM[1]	26.1	30
Fast R-CNN[2]	70.0	0.5
제안한 방법	63.4	45

DPM[1]방법은 먼저 화상에서 창문주사를 통하여 후보령역들을 얻고 분류기를 적용하여 대상을 검출하였다. Fast R-CNN[2]는 창문주사방법이 아니라 대상후보추출방법을 먼저 적용하고 추출된 후보들에 대하여 분류기를 적용한 R-CNN방법에서 분류단계의 속도를 올린 방법이다. 그러나 아직은 화상에서 후보령역제안을 생성하는데 2s정도 걸리는 선택적인 탐색에 의존하고있다.

표에서 보여준것처럼 제안된 방법이 Pascal VOC에 대한 검출에서 속도와 검출정확성의 측면에서 우수한 방법이라는것을 알수 있다.

맺 는 말

입력되는 화상을 1개의 중첩신경망을 리용하여 대상의 위치와 크기, 클라스를 결정하는 통합구조형대상검출방법을 제안하고 화상주사방식이 아니라 한번의 신경망예측결과를 리용하여 속도를 갱신하였다.

참 고 문 헌

[1] M. A. Sadeghi, D. Forsyth; Computer Vision–ECCV, 5, 6, 65, 2014.

[2] R. B. Girshick; CoRR, abs/1504.08083, 2, 5, 2015.

주체109(2020)년 11월 5일 원고접수

On the Study of Unified Architecture Object Detection Using CNN

Hong Kwang Chol

We present real-time object detection using single CNN, which predicts bounding boxes and class probabilities directly from whole images in one evaluation. Our unified architecture is extremely fast. Our base model deals with images in real-time at 45frames per second.

Keywords: neural network(NN), convolution neural network(CNN), object detection