

특징지도련결을 리용한 중첩신경망에 의한 한가지 대상검출방법

김동혁, 홍남운

우리는 특징지도련결을 리용한 중첩신경망에 의한 한가지 대상검출방법을 제기하고 실험을 통하여 효과성을 검증하였다.

선행연구[1]에서는 대상검출을 위한 SSD망을 제안하고 실험을 통하여 그 효과성을 검증하였다. 여기서는 기초망(VGG16)으로부터 2개의 층과 새로 추가하는 4개의 층으로부터 특징지도들을 추출하고 검출모형에 입력하여 대상분류와 대상령역회귀를 진행하였다.

선행연구[2]에서는 대상분류를 위한 새로운 오차함수를 제안하고 그를 리용하여 대상검출을 진행하고 그 효과성을 실험을 통하여 검증하였다.

선행연구[3]에서는 대상검출을 위한 NAS-FPN방법을 제기하고 실험을 통하여 그 효과성을 검증하였다. 그러나 신경망의 구조탐색은 매우 큰 계산량과 자원을 요구하는 부족점을 가진다.

론문에서는 여러 단계의 특징지도들을 련결하여 새로운 특징지도를 생성하고 그것을 리용하면서 선행연구[2]에서 제안한 손실함수를 적용하여 성능을 높이기 위한 중첩신경망을 제안하였다. SSD망[1]의 구조를 리용하면서 특징지도련결을 진행하여 보다 효과적인 특징지도를 대상검출에 리용하였다.

SSD망은 특징추출망으로부터 생성된 그물마디점들로부터 일정한 개수의 제안령역들을 생성하고 그 때 제안령역에서 종류별 대상출현에 대한 점수를 생성하는 정방향중첩신경망에 기초한 대상검출모형이다. 이 모형에서는 VGG16모형에서 화상분류를 위한 뒤부분의 전결합층을 제거하고 일부 층을 변경한 앞부분의 15개의 중첩층(5개의 층묶음)과 새로운 4개의 중첩층을 추가하여 특징추출망을 구성하였다. 그리고 특징추출망의 앞부분의 10번째와 15번째 중첩층으로부터 나오는 특징지도들과 뒤부분의 4개의 중첩층으로부터 나오는 특징지도들을 검출모형에 입력하여 대상분류 및 령역위치회귀를 진행하였다.

표 1. 특징추출망의 속성(입력화상의 해상도가 (300, 300)인 경우)

| 특징지도 | 해상도 | 걸음크기 | 본래화상에 대한 제 안령역의 크기비율 | 특징지도화소에서의 령역/개수 |
|--------|----------|------|-------------------------|--------------------|
| 특징지도 1 | (38, 38) | 8 | 0.1 | 4 |
| 특징지도 2 | (19, 19) | 16 | 0.2 | 6 |
| 특징지도 3 | (10, 10) | 32 | 0.375 | 6 |
| 특징지도 4 | (5, 5) | 64 | 0.55 | 6 |
| 특징지도 5 | (3, 3) | 100 | 0.725 | 4 |
| 특징지도 6 | (1, 1) | 300 | 0.9 | 4 |

추출되는 때 특징지도에 대한 제안령역생성은 특징추출망의 속성들을 리용하여 진행한다. 즉 특징지도 1에 대하여 표 1에서와 같이 때 특징지도화소에서의 제안령역의 크기 비율을 0.1로 설정(본래화상의 해상도가 (300, 300)일 때 특징지도 1에 대한 제안령역의

해상도는 (30, 30)임.)한다.

마찬가지로 특징지도 2로부터 특징지도 6에 대하여 매 특징화소에서의 제안령역의 크기비율을 0.2, 0.375, 0.55, 0.725, 0.9로 설정한다. 이에 기초하여 매 특징지도화소에 대하여 제안령역의 길이와 높이의 비율을 $\{1/2, 1, 2\}$ (혹은 $\{1/2, 1/3, 1, 2, 3\}$)로 정하여 3개 (혹은 5개)의 제안령역을 생성하고 식 (1)로부터 린접한 두 특징지도사이의 크기비율을 리용하여 1개의 제안령역을 추가적으로 생성하여 모두 4개(혹은 6개)의 제안령역들을 생성한다.

$$s'_k = \sqrt{s_k \cdot s_{k+1}} \quad (1)$$

여기서 s_k 는 매 특징지도에 대한 제안령역의 크기비율이다. 즉

$$s_1 = 0.1, s_2 = 0.2, s_3 = 0.375, s_4 = 0.55, s_5 = 0.725, s_6 = 0.9$$

이다. 또한 s'_6 의 계산을 위하여 $s_7 = 1.03$ 을 정의한다.

이로부터 모두 8 732개의 제안령역을 생성한다.

이제 특징지도 1에 대하여 고찰한다.

우에서 고찰한바와 같이 첫 특징지도는 10번째 중첩층으로부터 추출하며 걸음크기는 8이고 해상도는 (38, 38)이다.

본래화상에 대한 비율은 대략 $38/300 \approx 0.1$ 정도이지만 실지 접수마당의 크기를 확인해보면 (92, 92)이다. 10번째 중첩층이 포함된 4번째 층 묶음에는 3개의 중첩층이 포함되며 매 중첩층들의 걸음크기는 8이고 해상도는 (38, 38)로서 계산량을 크게 증가시키지 않으면서 련결가능성을 준다.

이 3개의 중첩층에 대하여 접수마당의 크기는 (60, 60), (72, 72), (92, 92)로서 실지 매 특징지도화소가 영향을 미치는 화상의 크기가 서로 다르며 특징지도수준도 중첩층들을 통과하면서 보다 높은 수준으로 변화된다.

론문에서는 이 3개의 중첩층으로부터 나오는 특징지도들을 련결하여 새로운 특징지도를 생성함으로써 작은 대상에 대한 검출에 보다 효과적하도록 한다. 일반적으로 중첩신경망에서 얇은 층에서는 대상의 위치정보를 잘 표현하고 보다 깊은 층일수록 대상의 의미정보를 잘 표현하는것으로 알려져있다.

론문에서는 대상의 위치정보를 잘 표현하는 제일 얇은 층의 특징지도를 새로운 특징지도로 변경함으로써 작은 대상에 대한 분류와 령역회귀의 정확성을 높인다.

특징지도련결단계에서는 3개의 중첩층이 출력하는 특징지도들을 련결하고 1×1 중첩층을 통과하여 특징지도를 생성하였다.

검출모형은 SSD구조에서와 같은 모형을 리용한다.

론문에서는 20개로 분류된 대상들에 대한 표식을 가지는 PASCAL VOC자료기지에 대하여 훈련과 시험을 진행하였다.

훈련은 PASCAL VOC2012자료기지의 훈련화상과 검증화상, PASCAL VOC2007자료기지의 훈련화상과 검증화상 22 136장에 대하여 진행하고 시험은 PASCAL VOC2007자료기지의 시험화상 4 952장에 대하여 진행하였다.

색변환, 화상의 축소와 확대를 진행한 다음 표본화, 수평뒤집기의 증식을 진행하였다.

모형무게결수는 특징추출망의 앞부분의 중첩층들에 대하여서는 ILSVRC CLS-LOC자료기지에서 훈련된 VGG16모형의 무게결수를 리용하고 특징추출망의 뒤부분의 중첩층들

과 특징지도런결의 중첩층의 무게결수에 대하여서는 선행연구[4]에서 제안한 무게결수를 리용하였다. 즉 $\mu=0$, $\sigma=\sqrt{2/n_i}$ 인 정규분포에 따르는 무게결수초기화를 진행하였다.

오차함수는 분류오차함수와 회귀오차함수의 무게합으로 구성하였다. 분류오차함수는 선행연구[2]에서 제안한 오차함수를 리용하고 회귀오차함수는 $smooth_{L1}$ 오차함수를 리용하였다.

$$L(x, c, l, g) = \frac{1}{N} (L_c(x, c) + \lambda L_l(x, l, g))$$

$$L_c(x, c) = - \sum_{i \in Pos} \sum_{j=1}^{N_g} x_{ij}^p (1 - \bar{c}_i^p)^r \log(\bar{c}_i^p) - \sum_{i \in Neg} (1 - \bar{c}_i^0)^r \log(\bar{c}_i^0), \quad \bar{c}_i^p = \frac{\exp(c_i^p)}{\sum_k \exp(c_i^k)}$$

$$L_l(x, j, g) = \sum_{i \in Pos} \sum_{j=1}^{N_g} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^p smooth_{L1}(l_i^m - \bar{g}_j^m)$$

$$\bar{g}_j^{cx} = \frac{g_j^{cx} - d_j^{cx}}{d_i^w}, \quad \bar{g}_j^{cy} = \frac{g_j^{cy} - d_i^{cy}}{d_i^h} \quad (2)$$

$$\bar{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right), \quad \bar{g}_j^h = \log\left(\frac{g_j^h}{d_i^h}\right)$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geq 1 \end{cases}$$

여기서 $x_{ij}^p \in \{1, 0\}$ 은 i 번째 제안령역이 j 번째 대상의 령역으로 될 때 1이고 그렇지 않으면 0이다. p 는 j 번째 대상의 종류이다. Pos 는 령역예측결과와 표식령역사이의 IOU가 0.5이상인 제안령역들의 모임이며 Neg 는 나머지 제안령역들의 모임중에서 우연적으로 Pos 에 속하는 제안령역개수의 3배로 취한다. N 은 Pos 에 속하는 제안령역개수이며 N_g 는 표식령역개수이다. c_i^p 는 i 번째 제안령역이 p 번째 종류의 대상으로 될 분류예측결과이다. $p=0$ 일 때 c_i^0 은 어떤 종류의 대상도 아닐 때의 분류예측결과이다.

γ 는 선행연구[2]에서 제안한 분류오차의 보조변수로서 2로 정한다.

l 은 령역예측결과이며 g 는 표식령역, d 는 제안령역을 의미한다.

$\{cx, cy, w, h\}$ 는 령역의 중심 및 길이, 너비를 의미한다.

훈련은 1개의 nVidia RTX 2080Ti영상가속장치에서 묶음크기는 32개, 전체 반복회수는 200으로 설정하여 40h동안 진행하였다. 학습률은 초기학습률을 10^{-3} 으로 설정하고 10^{-5} 까지 지수적으로 감소하도록 설정하였으며 Momentum최적화방법을 리용하였다.

훈련과정에 얻어지는 오차곡선은 그림과 같다.

오차곡선이 보여주는바와 같이 오차가 반복회수

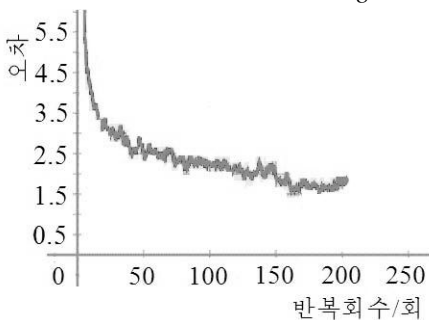


그림. 오차곡선

200까지 계속 수렴한다는것을 알수 있다. 실험결과 SSD모형과 제안한 모형의 대상검출에 리용하는 평가척도인 mAP와 FPS는 표 2와 같다.

표 2. mAP와 FPS

| 모형 | mAP | FPS |
|--------|-------|------|
| SSD | 0.772 | 59.3 |
| 제안한 모형 | 0.791 | 58.7 |

결과 제안한 모형은 SSD와 거의 같은 속도에서 더 높은 성능을 얻는다.

참 고 문 헌

- [1] W. Liu et al.; arXiv, vol.1512.02325, 2015.
- [2] T. Y. Lin et al.; arXiv, vol.1708.02002, 2017.
- [3] G. Ghiasi et al.; arXiv, vol.1904.07392, 2019.
- [4] K. He et al.; arXiv, vol.1502.01852, 2015.

주제109(2020)년 12월 5일 원고접수

An Object Detection Method by Convolutional Neural Network Using Concatenating Feature Maps

Kim Tong Hyok, Hong Nam Un

In this paper, we propose a kind of object detection method by convolutional neural network using concatenating feature maps and verify the efficiency on PASCAL VOC database.

Keywords: convolutional neural network, detection