

Exercícios Deep Learning

Aula 22

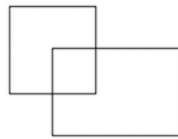
October 17, 2019

1 Janelas Deslizantes

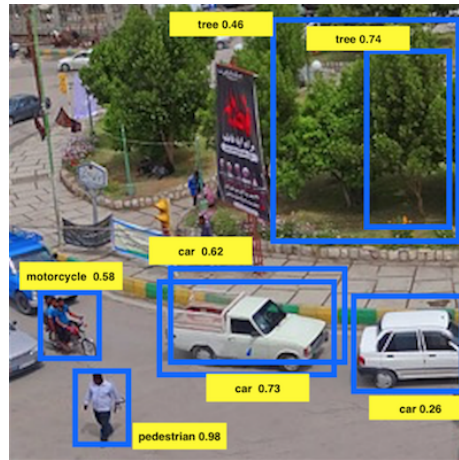
1- Considere um volume $2 \times 2 \times 2$, onde w_{ijc} é a célula (i, j) do canal c , 2 filtros 2×2 , onde F_{ijk} é o valor de cada célula (com k como identificador do filtro), e uma função de ativação $g(x)$. Encontre a forma fechada da saída dessa camada após a aplicação dos filtros. Explique como essa camada convolucional é equivalente à uma camada Fully Connected (indique o que seriam os nós, os pesos e suas dimensões).

2 Detecção de Objetos / YOLO

2- Qual é o valor da Interseção sobre a União (IoU) entre estas duas caixas? A caixa superior esquerda é 2×2 e a caixa inferior direita é 2×3 . A região sobreposta é 1×1 .



3- Suponha que você execute a non-max suppression nas caixas previstas na imagem abaixo. Os parâmetros usados são: caixas com probabilidade ≤ 0.4 são descartadas, e o limiar de IoU para decidir se duas caixas se sobrepõem é 0.5. Indique quantas e quais caixas permanecerão.



- 4- Repita o exercício anterior com probabilidade de descarte igual a 0.6.
- 5- Cite 2 situações nas quais o algoritmo de non-max suppression é falho.
- 6- No algoritmo YOLO, em tempo de treinamento, como ocorre a detecção do objeto caso ele esteja em mais de uma célula do grid da imagem?
- 7- Suponha que você esteja usando o YOLO com uma grade 19×19 , em um problema de detecção com 20 classes e 5 anchor boxes. Qual é a dimensão do volume de saída da rede?
- 8- Sabendo que o Yolo é consideravelmente mais rápido que a R-CNN, Fast R-CNN e a Faster R-CNN, por que ele não é o mais utilizado em todos os cenários? Quais são os seus problemas?
- 9- Por que o YOLO não consegue detectar diferentes objetos muito próximos?
- 10- Qual a vantagem de se utilizar múltiplas Anchor Boxes em detecção de objetos?
- 11- Cite duas formas de escolher Anchor Boxes e qual é a vantagem de cada uma delas.

Solução

1- Para cada ω_k de saída: $\omega_k = g(\sum_{i,j,c=1}^2 w_{ijc} F_{ijck} + b_k)$, onde ω_k são os nós de saída com dimensão 2, w_{ijc} são os nós de entrada com dimensão 8, F_{ijck} são os pesos com dimensão $8 \times 2 = 16$ e b_k é o bias com dimensão 2. Portanto essa camada convolucional é equivalente a uma camada fully connected com 8 nós de entrada e dois de saída.

$$\mathbf{2-} \text{IoU} = \frac{I}{U} = 1/9, I = b_1 \cap b_2 = 1, U = b_1 \cup b_2 = 9$$

3- 5 Caixas.

Algoritmo do Non max-supression:

- Descarte todas as BB com probabilidade menor ou igual a 0.4
- Enquanto ainda houver BB não analisadas: a) Escolha a BB de maior probabilidade; b) Remova qualquer BB que tenha IoU maior que 0.5 com a BB escolhida em (a).

4- 3 Caixas.

5- Quando a bounding box de maior probabilidade não é a mais adequada; Objetos distintos estão muito perto uns dos outros e Falsos positivos.

6- Apenas uma célula - a que contém o centro/ponto médio de um objeto - é responsável por detectar esse objeto, já a anchor box vai além dos limites do grid.

7- Dimensão final: $19 \times 19 \times (5 \times 25)$, onde: 19×19 para cada quadrado da grade analisada, 5 é a quantidade de anchor boxes, e 25 se refere a 20 classes mais 5 dimensões para p_c, b_x, b_y, b_h, h_w .

8- Alguns problemas do YOLO são: Dificuldade de detectar objetos pequenos devido à limitação do grid, detecção de aspectos com proporções distorcidas, ponderar de forma igual erro em bounding boxes grandes e pequenas.

9- No YOLO, ainda que depois de realizar a divisão da imagem em um grid sejam preditas B bounding boxes cada uma com seu score de confiança, se a interseção entre essas caixas não for muito pequena, apenas um objeto será detectado.

10- O uso de múltiplas anchor boxes permite que uma rede detecte vários objetos, objetos de diferentes escalas e objetos sobrepostos.

11- O primeiro modo é manualmente, a vantagem é que é um método simples. O segundo modo é utilizar o algoritmo k-means, a vantagem é que as caixas serão mais adequadas aos dados.