

Exercícios Deep Learning

Segmentação Semântica

- 1) O que é segmentação semântica?
- 2) Por que é dito que a tarefa de segmentação semântica é uma tarefa de predição densa enquanto a detecção de objetos é uma tarefa de predição esparsa?
- 3) Quais são as diferenças entre as tarefas de: classificação de imagens, detecção de objetos, segmentação semântica e segmentação de instâncias?
- 4) Quais são algumas aplicações da segmentação semântica?
- 5) Considere uma arquitetura treinada para fazer segmentação semântica em 2 classes. Explique como é possível criar um mapa de probabilidades onde cada pixel na sua imagem de saída é uma representação da probabilidade do pixel pertencer a uma determinada classe.
- 6) Descreva como é possível transformar uma arquitetura de classificação de imagens em uma de segmentação semântica. Qual é uma possível desvantagem dessa abordagem?
- 7) Por que uma rede neural Fully Connected não funciona em tarefas de segmentação semântica/de instância?
- 8) Quais são algumas arquiteturas comuns usadas em modelos de segmentação semântica?
- 9) Em uma Fully Convolutional Network, qual é o objetivo do downsampling e upsampling?
- 10) Considere o seguinte volume de entrada:

$$\begin{bmatrix} 5 & 6 \\ 3 & 2 \end{bmatrix}$$

Faça o upsampling do volume para uma saída 4×4 de acordo com cada um dos métodos:

- a) Nearest Neighbor
- b) Bed of Nails ("Cama de Pregos")
- c) Max unpooling, considere que o max pooling correspondente foi feito na seguinte matriz:

$$\begin{bmatrix} 3 & 1 & 1 & 4 \\ 2 & 1 & 3 & 2 \\ 3 & 7 & 1 & 7 \\ 3 & 2 & 2 & 4 \end{bmatrix}$$

d) Faça o upsampling para uma saída 3×3 através da convolução transposta com padding=0 e stride=1, considerando o seguinte filtro:

$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$$

Solução

- 1) A segmentação semântica é uma técnica que envolve atribuir um rótulo (classe) a cada pixel em uma imagem, permitindo um entendimento mais detalhado da cena retratada na imagem.
- 2) Em segmentação semântica, cada pixel da imagem tem um rótulo que deve ser predito pelo classificador. Em detecção de objetos, a maior parte da imagem pode não ser rotulada, pois estamos interessados somente nos rótulos dos objetos capturados pelas bounding boxes.
- 3) Classificação de imagens: Dado uma imagem, a rede deve gerar um rótulo para ela.
Detecção de objetos: Dada uma imagem a rede deve detectar os objetos, localizá-los numa bounding box e prever a classe do objeto.
Segmentação semântica: Dada uma imagem cada pixel deve ser classificado em uma classe (que pode ou não ser um objeto).
Segmentação de instâncias: Dada uma imagem, localizar os objetos e identificar cada pixel pertencente a cada instância do objeto.
- 4) A segmentação semântica é usada em uma variedade de aplicações, incluindo carros autônomos, imagens médicas e análise de imagens de satélite.
- 5) Dado o treinamento da arquitetura, durante o forward-propagation de uma imagem de teste, basta utilizar a saída da função softmax, que retorna um vetor de probabilidades entre as possíveis classes treinadas, e fazer o mapeamento direto dos intervalos de [0,1] em [0,255] para cada pixel. O novo mapeamento será visto como um tom de cinza dentro da imagem de saída.
- 6) Dado uma imagem de entrada para classificação pixel a pixel, basta fazer pequenos recortes de tamanho ímpar (pois é necessário um pixel central). Ao passar esse recorte pela rede, teremos uma classificação que será usada no pixel central. Repetiremos o processo para todos os pixels na imagem e assim teremos uma saída para a tarefa de classificação semântica. Como o processo deve ser repetido para cada pixel da imagem, o custo computacional dessa abordagem é muito alto.
- 7) Ao transformar os pixels de uma imagem em nós de uma rede neural, a orientação espacial de cada um em relação aos outros é perdida, o que é necessário para realizar a segmentação semântica/de instância.
- 8) Arquiteturas comuns usadas em modelos de segmentação semântica incluem Redes Totalmente Convolucionais (FCN), SegNet e DeconvNets.
- 9) O Downsampling diminui custo de computação e memória pois as camadas intermediárias terão tamanho menor. Já o Upsampling é necessário para alinhar a previsão da rede com os pixels da imagem.

10)

$$\text{a)} \begin{bmatrix} 5 & 5 & 6 & 6 \\ 5 & 5 & 6 & 6 \\ 3 & 3 & 2 & 2 \\ 3 & 3 & 2 & 2 \end{bmatrix} \quad \text{b)} \begin{bmatrix} 5 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ 3 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{c)} \begin{bmatrix} 5 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{d)} \begin{bmatrix} 0 & 5 & 6 \\ 10 & 30 & 20 \\ 6 & 13 & 6 \end{bmatrix}$$