

## Problem 1

از لحاظ تعداد پارامتر تعداد ~~شکلهای~~ <sup>شکلهای</sup> CNN ندارند و شبکه های FC  
 قابلیت استخراج local feature ها را بهتر دارد. شبکه دوم بهتر است  
 برای classification task

تعداد پارامتری FC  $\rightarrow 5(100+1) = 505$   
 تعداد پارامتر CNN  $\rightarrow 5(10 \times 10 + 1) = 505$

## Problem 2

Relu max pooling  $\rightarrow$  پارامتر ندارد

لایه اول  $\rightarrow$   $\begin{matrix} \text{bias} \\ \uparrow \\ 16(5 \times 5 \times 3 + 1) = 1216 \end{matrix}$   
 تعداد پارامتر =  $\frac{128 \times 128}{3} = 128 \times 128 \times \frac{1}{3}$   
 Conv  $\rightarrow$   $\frac{128 + 1 - 5}{1} = 128$   
 Pooling  $\rightarrow$   $\frac{128 - 2}{2} + 1 = 64$

لایه دوم  $\rightarrow$   $\begin{matrix} \text{bias} \\ \uparrow \\ 16(5 \times 5 \times 16 + 1) = 6416 \end{matrix}$   
 تعداد پارامتر =  $\frac{64 \times 64}{16} = 64 \times 64 \times \frac{1}{16}$

Conv  $\rightarrow$  64

Pool  $\rightarrow$  32

لایه سوم  $\rightarrow$   $\begin{matrix} \text{bias} \\ \uparrow \\ 16(32 \times 32 \times 16 + 1) = 14048 \end{matrix}$   
 تعداد پارامتر =  $32 \times 32 \times 16$

Conv  $\rightarrow$  32

Pool  $\rightarrow$  16

تعداد پارامتر =  $16 \times 16 \times 16$  و  $\begin{matrix} \text{تعداد کل} \\ \text{پارامترها} \end{matrix} = 14048$

$\begin{matrix} \text{bias} \\ \uparrow \\ (14048) + (16 \times 16 \times 16 + 1) \times 10 = 55018 \end{matrix}$

↑  
CNN-Net

d)

$$n_{out} = \left\lfloor \frac{n + 2p - k}{s} \right\rfloor + 1$$

$$j_{out} = j_{in} \times s \quad \therefore Y_{out} = f_{in} + (k-1) \times j_{in}$$

الاول  $\rightarrow$  conv  $\rightarrow n_{out} = 128, j = 1, s = 1, k = 5$   
 $\Rightarrow Y_{out} = 1 + 4 = 5$

max pool  $\rightarrow n_{out} = 64, j_{out} = 2, k = 2$   
 $Y_{out} = 5 + 1 = 6$

الثاني  $\rightarrow$  conv  $\rightarrow n_{out} = 64, j = 2, k = 5$   
 $\rightarrow Y_{out} = 6 + 8 = 14$

max pool  $\rightarrow n_{out} = 32, j_{out} = 4, k = 2$   
 $Y_{out} = 14 + 2 = 16$

الثالث  $\rightarrow$  conv  $\rightarrow n_{out} = 32, j_{out} = 4, k = 5$   
 $Y_{out} = 1 + 16 = 17$

max pool  $\rightarrow n_{out} = 16, j_{out} = 8, k = 2$   
 $Y_{out} = 17 + 1 = 18$



### Problem 3

#### Part 1

(a) Encoder section ← کاهش بعد را انجام داده و ویژگی‌های مهم را استخراج می‌کند با استفاده از CNN

Bridge ← پیوند بین Encoder و Decoder و کار آن انتقال ویژگی‌هاست که شامل skip connections و Concatenation می‌باشد

Decoder section ← این قسمت با استفاده از اطلاعات bridge و upsampling را انجام داده و سایر up conv

تصاویر را بازسازی می‌کند

مسافتار شکل برای بهبود مسافتار pixel-level است که از دو مسیر مختلف شکل گرفته localization

(b)

1. preserving Spatial information

2. Gradient Flow and Training Stability

3. Enhancing Feature Reuse

4. combine local and global information effectively

(c) زیرا این نوع اتصالات به فرآیند انتقال اطلاعات مکانی را حفظ کرده و در فرآیند بازسازی از این نوع اطلاعات بهره می‌برد و

بابت افزایش دقت می‌باشد → برای segmentation

## Part 2

a)  $256 \times 256 \rightsquigarrow 254 \times 254 \rightsquigarrow 252 \times 252$

$\rightsquigarrow 126 \times 126 \rightarrow 124 \times 124 \rightsquigarrow 122 \times 122$

$\rightsquigarrow 61 \times 61 \rightsquigarrow 59 \times 59 \rightsquigarrow 57 \times 57$

$\rightsquigarrow 28 \times 28 \rightsquigarrow 26 \times 26 \rightsquigarrow 24 \times 24$

$\rightsquigarrow 12 \times 12 \rightsquigarrow 10 \times 10 \rightsquigarrow 8 \times 8$

$8 \times 8 \times C \leftarrow \text{سایر ویژگی}$   
 $\downarrow$   
 $1024$

b)  $128 \times 128$  فیلتر  
 $128 (3 \times 3 \times 64 + 1) + 128 (3 \times 3 \times 128 + 1) = 221440$

## Part 3

در ResNet و skip connection به صورت این صورت است که  
 $block$

a)  $H(x) = F(x) + x$

در DenseNet و از تمام لایه‌های قبل خود استفاده می‌کند  
 $block$

$$H(x) = H_e([x, H_1, \dots, H_{L-1}])$$

به این دلیل و شبکه‌های DenseNet قابلیت به هم پذیری بهتر دارند و باعث کاهش gradient vanishing می‌شود.

در شبکه‌های DenseNet و بهنگام یادگیری بهتر مدیریت می‌شود چون مسیرهای کوتاه‌تری وجود دارد در backpropagation برای Feature map ها هر لایه وجود دارد که باعث بهبود یادگیری آن‌ها می‌شود.

مزیت استفاده از ویژگی‌های لایه قبل در این است که شبکه در هر لایه ویژگی‌های جدیدتری را یاد می‌گیرد که این عمل باعث افزایش دقت و بازآموزش می‌شود.



Part 4

a)  $\underline{64 + 128 = 192}$

b)  $24 \times 3 + 32 = 104$