

Image Segmentation | U-Net

Hobot | Deep Learning | Zahra Amini

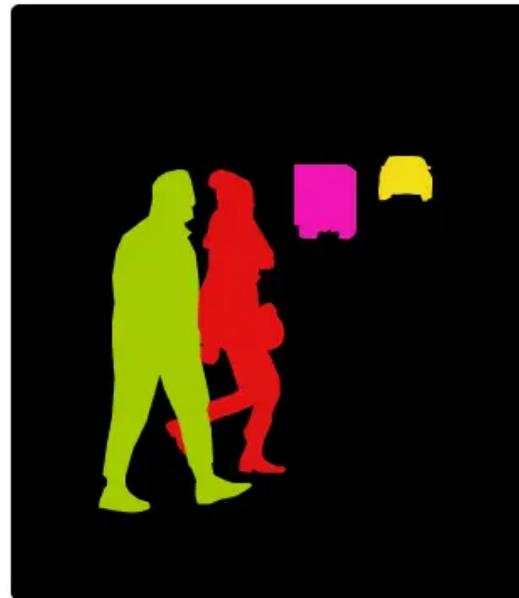
Telegram: @zahraamini_ai & Instagram:@zahraamini_ai & LinkedIn: @zahraamini-ai

<https://zil.ink/zahraamini>

Types of Image Segmentation



SEMANTIC IMAGE
SEGMENTATION



INSTANCE
SEGMENTATION



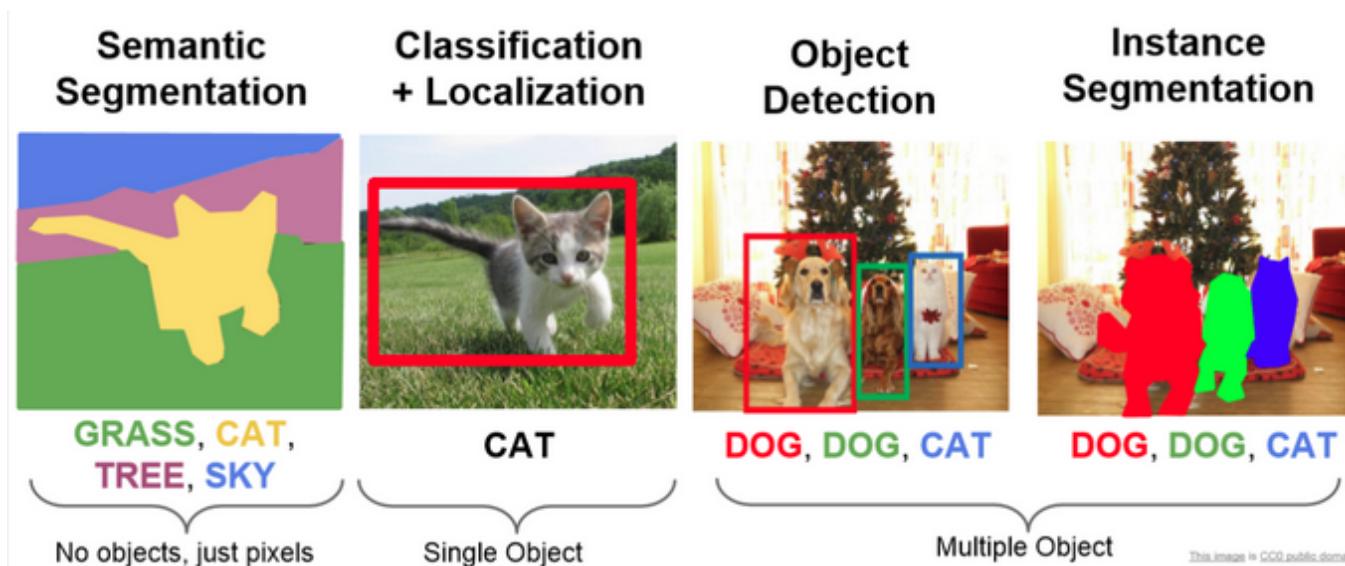
PANOPTIC
SEGMENTATION

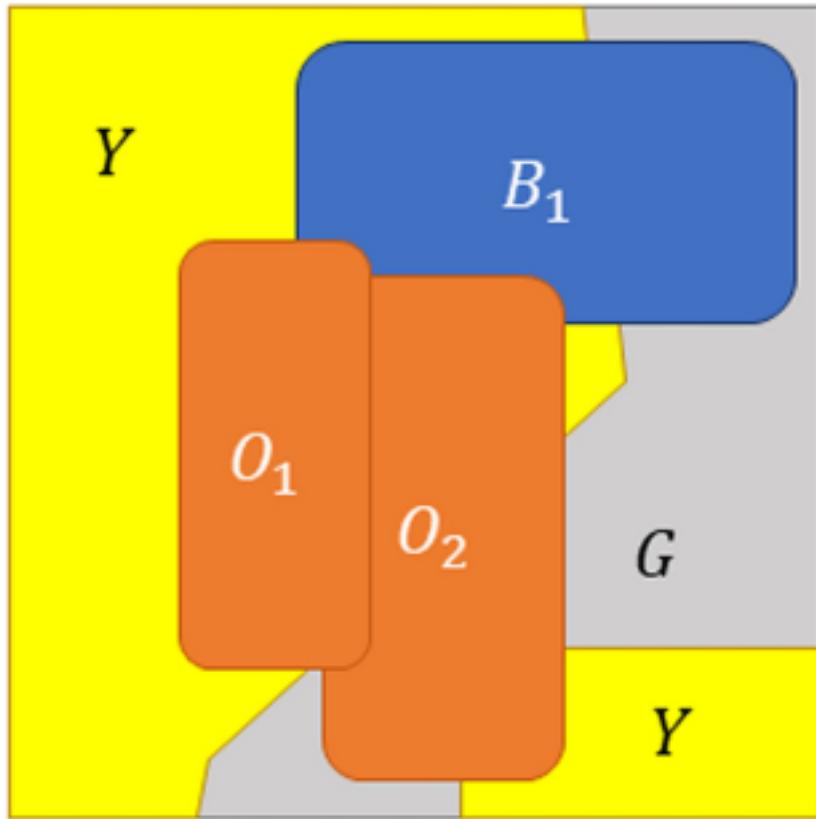
1. سگمنتیشن معنایی (Semantic Segmentation):

در این نوع سگمنتیشن، هر پیکسل تصویر به یکی از دسته‌های از پیش تعریف شده اختصاص داده می‌شود. به عبارت دیگر، تمام پیکسل‌هایی که به یک کلاس خاص (مانند انسان، ماشین، درخت و غیره) تعلق دارند، به یک رنگ مشابه رنگ‌آمیزی می‌شوند. در این روش، تفاوت بین اشیاء مشابه در تصویر مشخص نمی‌شود؛ یعنی اگر دو انسان در تصویر وجود داشته باشند، هر دو با یک رنگ یکسان نشان داده می‌شوند.

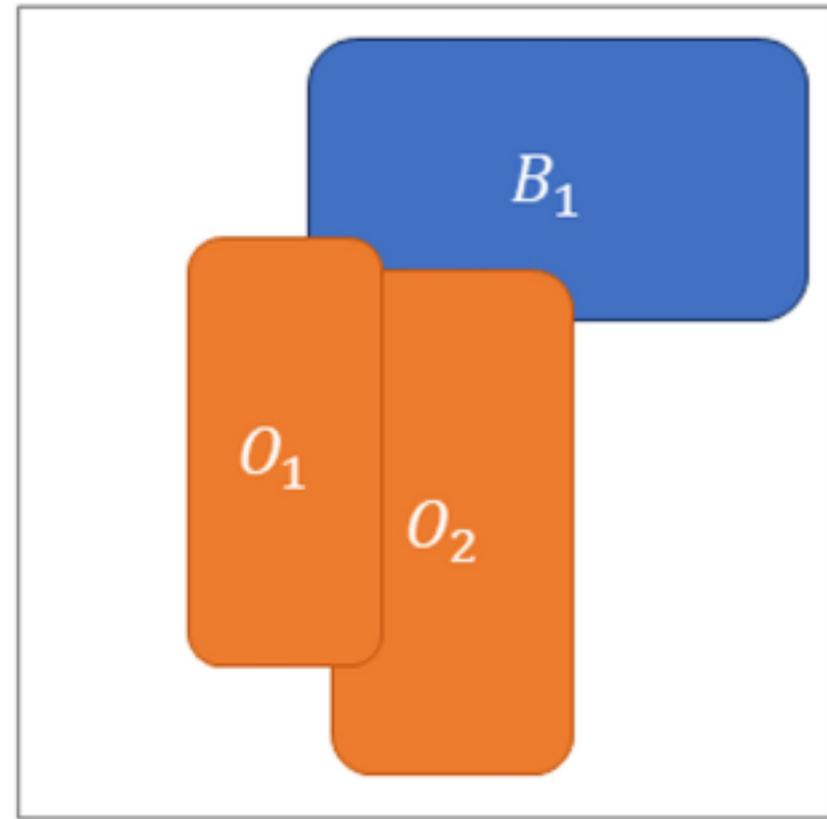
2. سگمنتیشن نمونه‌ای (Instance Segmentation):

این روش مشابه سگمنتیشن معنایی است با این تفاوت که در آن، علاوه بر تشخیص کلاس‌های اشیاء، هر نمونه از یک کلاس به طور جداگانه مشخص می‌شود. به عبارت دیگر، هر پیکسل نه تنها به یک کلاس خاص تعلق دارد، بلکه به یک نمونه خاص از آن کلاس نیز مرتبط است. برای مثال، اگر دو انسان در تصویر وجود داشته باشند، هر کدام با رنگ متفاوتی نشان داده می‌شوند تا تفاوت آنها مشخص شود.





Panoptic Segmentation

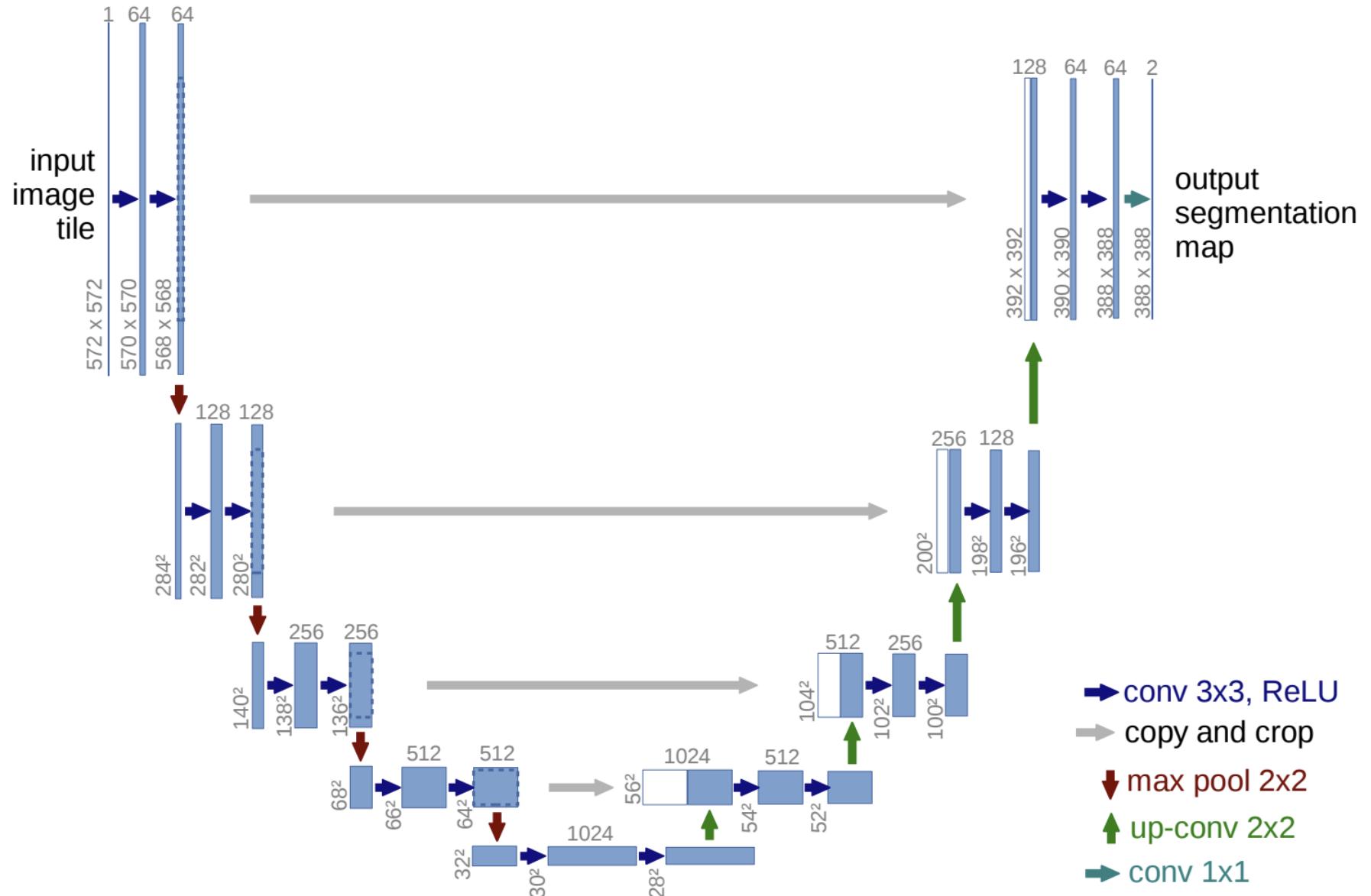


Instance Segmentation
and classification

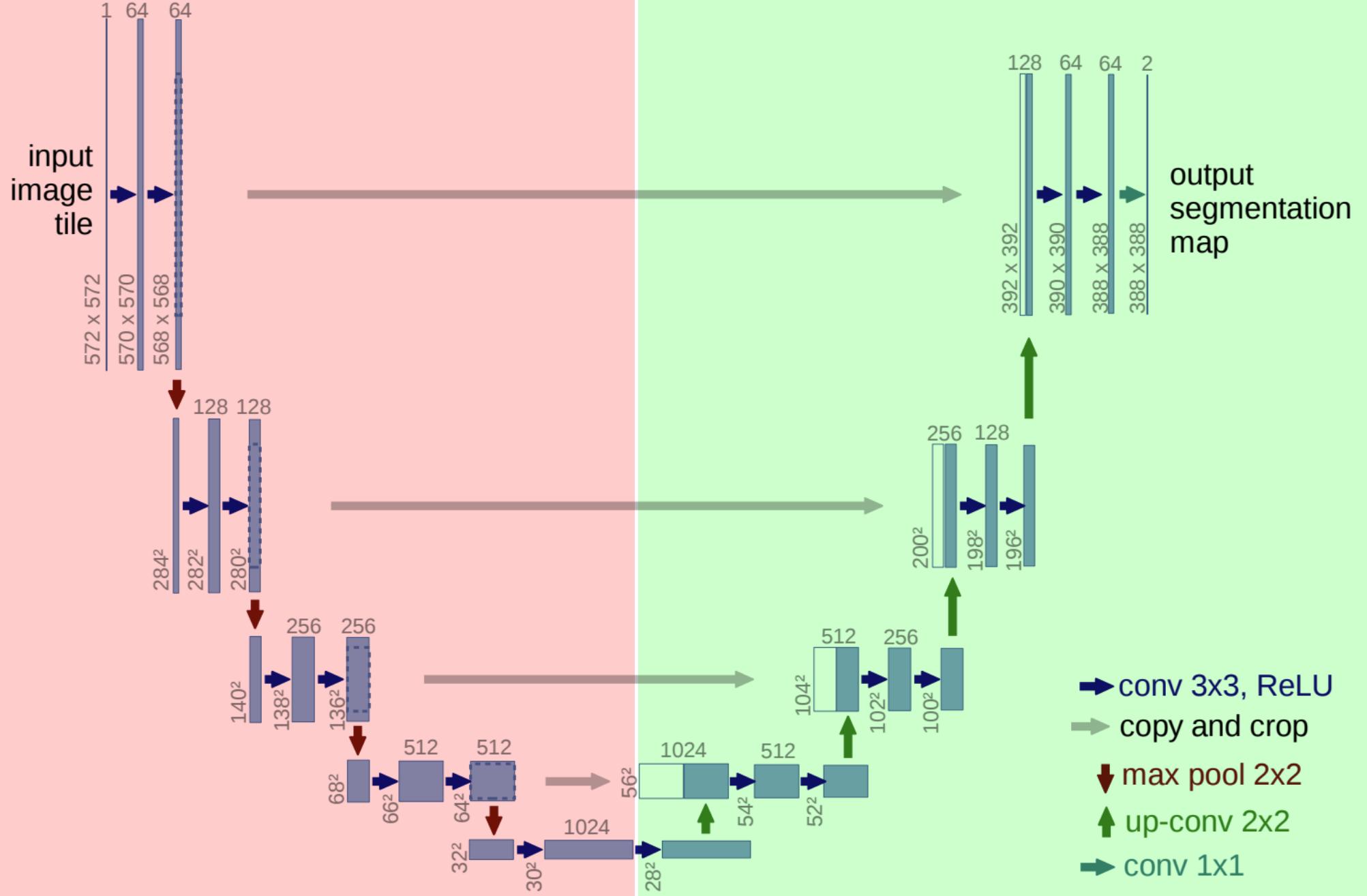
سگمنتیشن اینستنس فقط بر شناسایی و دسته‌بندی اشیاء خاص در تصویر تمرکز دارد و به پس زمینه توجهی نمی‌کند
سگمنتیشن پانوراما به تمام پیکسل‌های تصویر توجه می‌کند و نه تنها اشیاء خاص، بلکه تمام بخش‌های تصویر را نیز دسته‌بندی می‌کند

به این ترتیب، سگمنتیشن پانوراما دید جامع‌تری از تصویر ارائه می‌دهد و برای کاربردهایی که نیاز به اطلاعات کامل‌تری از تصویر دارند، مفید‌تر است

U-Net



یکی از شبکه‌های عصبی عمیق است که به طور ویژه برای سگمنتیشن تصاویر طراحی شده است. این مدل در ابتدا برای کاربردهای پزشکی توسعه داده شد، به ویژه برای سگمنتیشن تصاویر میکروسکوپی، اما امروزه در بسیاری از زمینه‌های دیگر نیز استفاده می‌شود. ساختار این شبکه به گونه‌ای است که ترکیبی از ویژگی‌های محلی و کلی را در تصویر بررسی می‌کند.



بخش فشرده‌سازی (Encoder):

این بخش شامل چندین لایه کانولوشنی (Convolutional Layers) است که به تدریج ابعاد تصویر ورودی را کاهش می‌دهند.

در هر مرحله، تعداد فیلترها افزایش می‌یابد تا ویژگی‌های پیچیده‌تری از تصویر استخراج شود.

در این بخش، از عملیات Max Pooling برای کاهش ابعاد استفاده می‌شود. این کار باعث کاهش اندازه فضایی تصویر می‌شود ولی اطلاعات مهمی از ویژگی‌های تصویر حفظ می‌گردد.

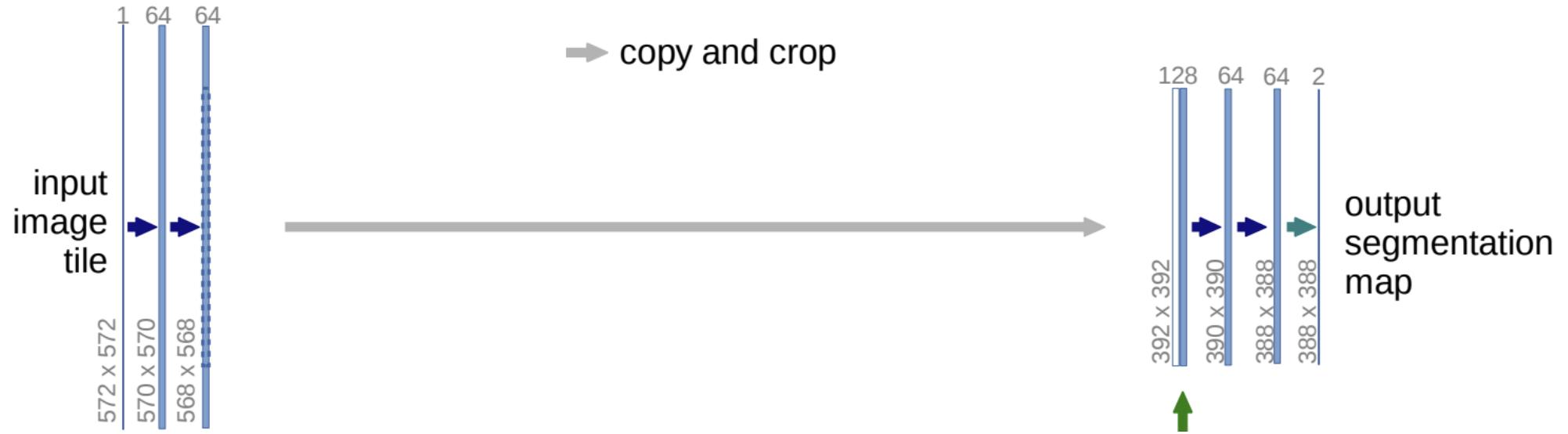
بخش گسترش (Decoder):

این بخش وظیفه بازیابی ابعاد اولیه تصویر و تولید سگمنت‌های نهایی را دارد.

از لایه‌های Deconvolution یا Upsampling برای افزایش ابعاد تصویر استفاده می‌شود.

در هر مرحله، لایه‌های خروجی از بخش فشرده‌سازی به لایه‌های متناظر در بخش گسترش متصل می‌شوند (Skip Connections). این اتصالات مستقیم باعث می‌شود که اطلاعات فضایی اصلی که در لایه‌های اولیه از دست رفته است، دوباره بازیابی شود.

Copy & Crop ---> Skip Connections



در U-Net، هر زمان که یک عملیات کانولوشن و سپس یک عملیات Max Pooling در بخش Encoder انجام می‌شود، خروجی این عملیات‌ها به یک لایه متناظر در بخش Decoder متصل می‌شود.

این اتصال به این صورت است که ویژگی‌های فضایی و اطلاعات دقیق‌تر که در مراحل ابتدایی در بخش Encoder استخراج شده‌اند، به مراحل متناظر در بخش Decoder افزوده می‌شوند.

این عمل باعث می‌شود که بخش Decoder، اطلاعات دقیق‌تری از ساختارهای تصویر داشته باشد و بتواند با جزئیات بیشتری تصویر نهایی را بازسازی کند.



Step 1: Input Image

$$\text{Input Image} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

Step 2: Convolution (Conv 3x3, ReLU)

$$\text{Filter} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{Conv Output} = \begin{bmatrix} 30 & 35 \\ 50 & 55 \end{bmatrix}$$

Top-left corner (output position (1,1)):

$$\text{Conv} = (1 \times 1 + 2 \times 0 + 3 \times -1) + (5 \times 1 + 6 \times 0 + 7 \times -1) + (9 \times 1 + 10 \times 0 + 11 \times -1) = -2$$

Step 3: Max Pooling (2x2)

Max Pool Output = [55]

Step 4: Upsampling with ConvTranspose

$$\text{Filter} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$55 \times \text{Filter} = \begin{bmatrix} 55 & 0 \\ 0 & 55 \end{bmatrix}$$

$$\text{Final Output} = \begin{bmatrix} 55 & 0 & 0 & 0 \\ 0 & 55 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

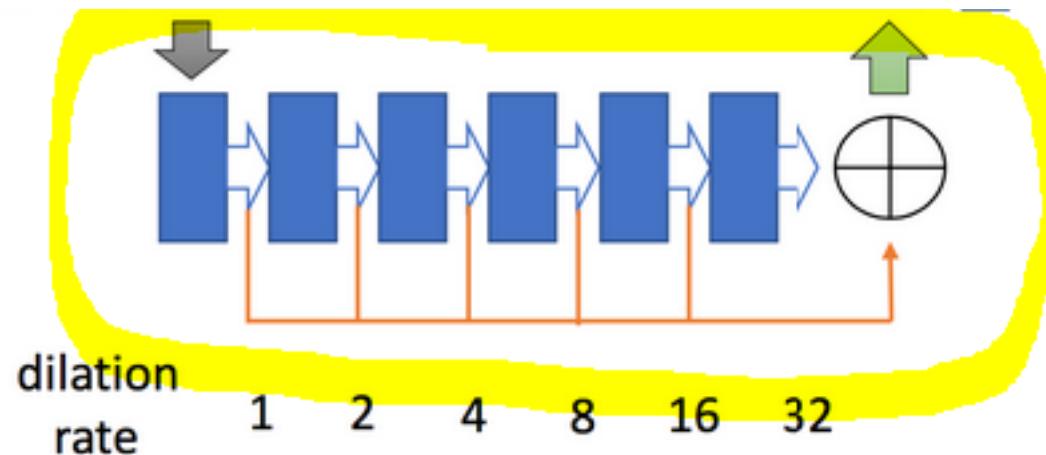
Step 5: Skip Connections

Suppose the feature map from the encoder at the corresponding stage was:

$$\text{Encoder Feature Map} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\text{Concatenated Output} = \begin{bmatrix} 55 + 1 & 0 + 1 & 0 + 1 & 0 + 1 \\ 0 + 1 & 55 + 1 & 0 + 1 & 0 + 1 \\ 0 + 1 & 0 + 1 & 0 + 1 & 0 + 1 \\ 0 + 1 & 0 + 1 & 0 + 1 & 0 + 1 \end{bmatrix} = \begin{bmatrix} 56 & 1 & 1 & 1 \\ 1 & 56 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

BottleNeck



این بخش فشرده‌ترین سطح شبکه است که در آن ابعاد تصویر به کمترین مقدار می‌رسد، اما تعداد ویژگی‌ها (فیلترها) در بیشترین مقدار است

کانولوشن با نرخ اتساع (Dilated Convolution)

در این بخش، چندین لایه کانولوشن با فیلترهای 3×3 و با نرخ اتساع مختلف اعمال می‌شود. نرخ اتساع به این معناست که فیلترها به گونه‌ای اعمال می‌شوند که هر کدام تعداد مشخصی پیکسل را نادیده می‌گیرند تا بتوانند اطلاعات را از مقیاس‌های مختلف در تصویر استخراج کنند.

در تصویر، نرخ اتساع به صورت 1, 2, 4, 8, 16, 32 نشان داده شده است. به عنوان مثال، نرخ اتساع 1 به معنای یک کانولوشن عادی است، در حالی که نرخ اتساع 2 به معنای این است که بین هر دو پیکسل که فیلتر اعمال می‌شود، یک پیکسل نادیده گرفته می‌شود.

Dilated Convolution

یک نوع عملیات کانولوشن است که در آن بین المان‌های فیلتر فاصله‌هایی ایجاد می‌شود و نقش مهمی ایفا می‌کند، به ویژه زمانی که هدف، افزایش دامنه اطلاعات مکانی بدون افزایش بار محاسباتی باشد

ابعاد فضایی تصویر را افزایش نمی‌دهد، بلکه میدان دید مؤثر را گسترش می‌دهد و در عین حال ویژگی‌های مهم را حفظ می‌کند

افزایش میدان دید: با استفاده از فواصل بین المان‌های فیلتر، کانولوشن گسترشی می‌تواند بخش‌های وسیع‌تری از تصویر را تحت پوشش قرار دهد بدون اینکه نیاز به افزایش ابعاد فیلتر یا کاهش وضوح فضایی داشته باشد

حفظ ویژگی‌ها: کانولوشن گسترشی این امکان را فراهم می‌کند که ویژگی‌های مهم از بخش‌های مختلف و بافت‌های گسترده‌تر تصویر بدون از دست دادن اطلاعات کلیدی حفظ شوند

CNN

$$\text{Input Matrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{bmatrix} \quad \text{Filter} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

$$\text{Result} = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 7 & 8 \\ 11 & 12 & 13 \end{bmatrix} \odot \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} -6 & -6 & -6 \\ -6 & -6 & -6 \\ -6 & -6 & -6 \end{bmatrix}$$

Dilated Convolution with Dilation Rate 2

$$\text{Input Matrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & - & 3 & - & 5 \\ - & - & - & - & - \\ 11 & - & 13 & - & 15 \\ - & - & - & - & - \\ 21 & - & 23 & - & 25 \end{bmatrix}$$

$$\begin{bmatrix} 1 & - & 3 & - & 5 \\ - & - & - & - & - \\ 11 & - & 13 & - & 15 \\ - & - & - & - & - \\ 21 & - & 23 & - & 25 \end{bmatrix}$$

$$\text{Dilated Result} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 11 & 0 & 13 \end{bmatrix} \odot \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

$$\text{Dilated Result}(1,1) = (1 \times 1) + (0 \times 0) + (3 \times -1) + (0 \times 1) + (0 \times 0) + (0 \times -1) + (11 \times 1) + (0 \times 0) + (13 \times -1) = 1 - 3 + 11 - 13 = -4$$

$$\text{Output (Dilation Rate} = 2) = \begin{bmatrix} -4 & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix}$$

UpSampling ---> ConvTranspose

ConvTranspose to Increase the Size of a 2×2 Matrix to 4×4

$$X = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

1. Create an Intermediate Matrix with Padding and Stride

$$\text{Intermediate Output} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 3 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2. Apply the Filter to the Intermediate Matrix

$$F \times 1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times 1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{Output after first step} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$F \times 2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times 2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{Output after second step} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$F \times 3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times 3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{Output after third step} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{bmatrix}$$

$$F \times 4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times 4 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \quad \text{Final Output} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 4 & 0 \\ 0 & 3 & 0 & 4 \end{bmatrix}$$