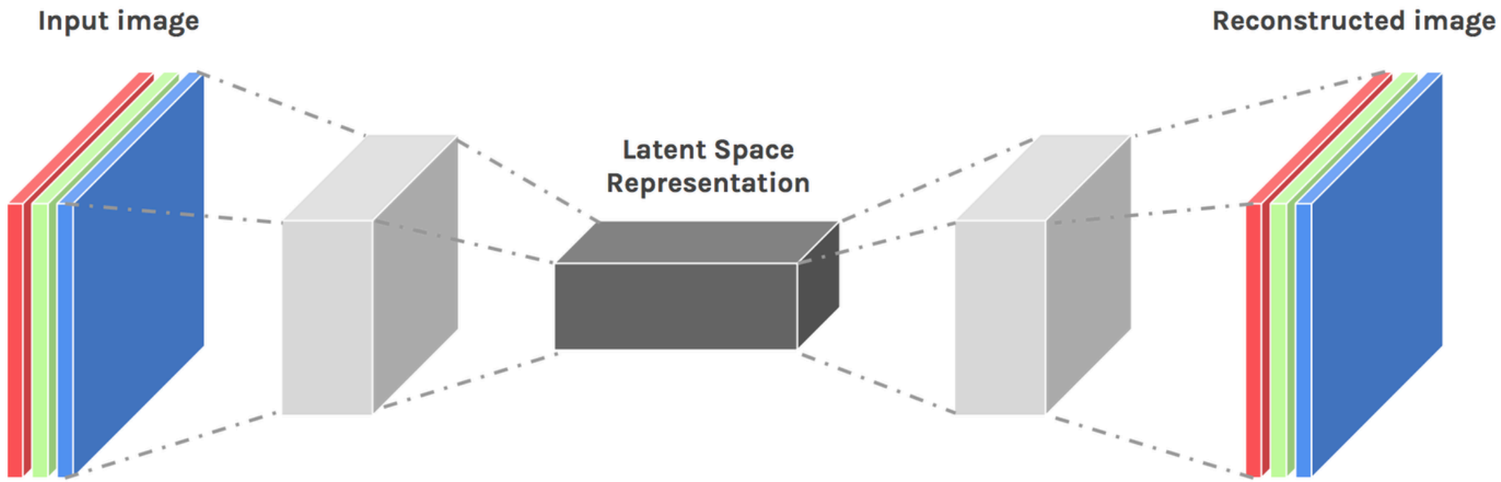
ENCODER AND DECODER CONVOLUTIONAL NETWORKS

Astalus Adrian

KEY COMPONENTS

The decoder part of the network takes the input and converts it into an encoded form that retains the most important and relevant information.

The decoder then takes the encoded representation from the latent space and reconstructs it into the desired output shape.



USAGE

Encoder-decoder networks are highly versatile architectures that are primarily used in tasks where the goal is to map an input to an output, often with different sizes or spatial dimensions.'

Some of these tasks include:

- Image Denoising
- Image Super-Resolution
- Image Segmentation
- Image-to-Image Translation

IMAGE DENOISING

Image denoising consists of improving an image's clarity by removing the information that makes the image less clear, also known as noise.

Some of the more popular methods for image denoising are **denoising autoencoders** and **U-Net**.

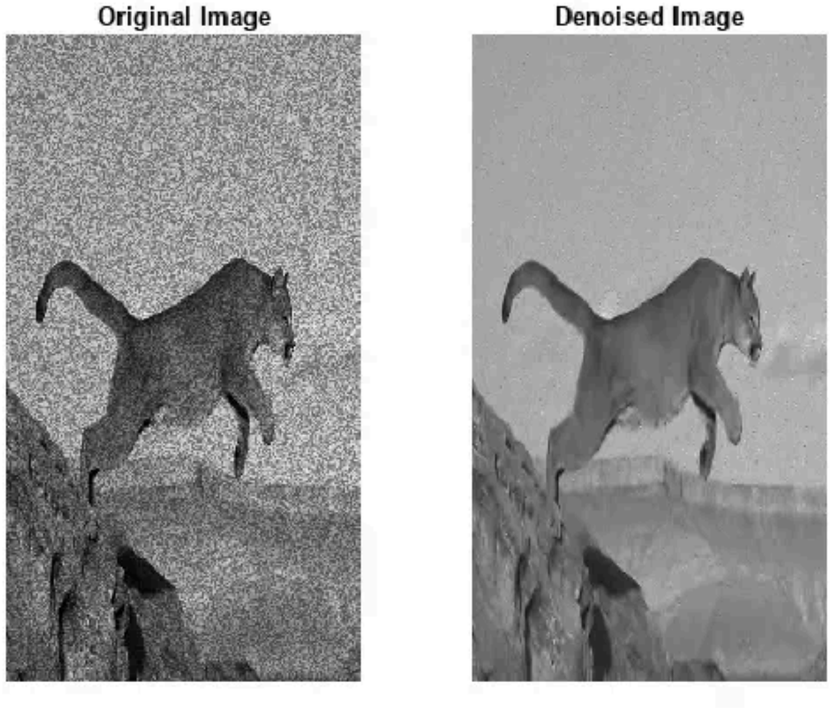
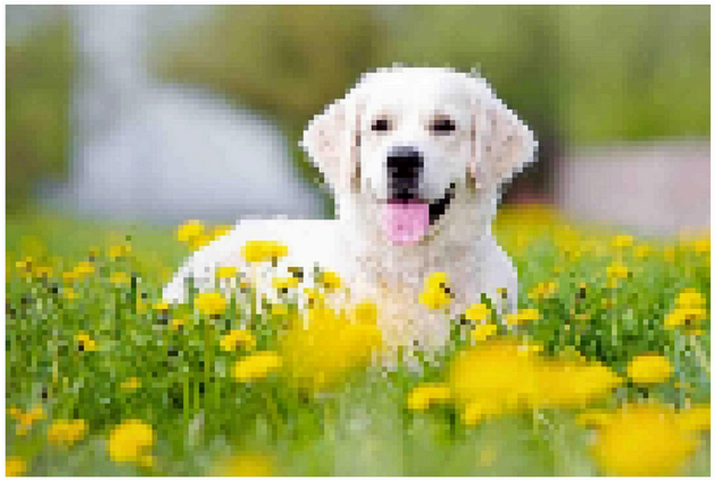
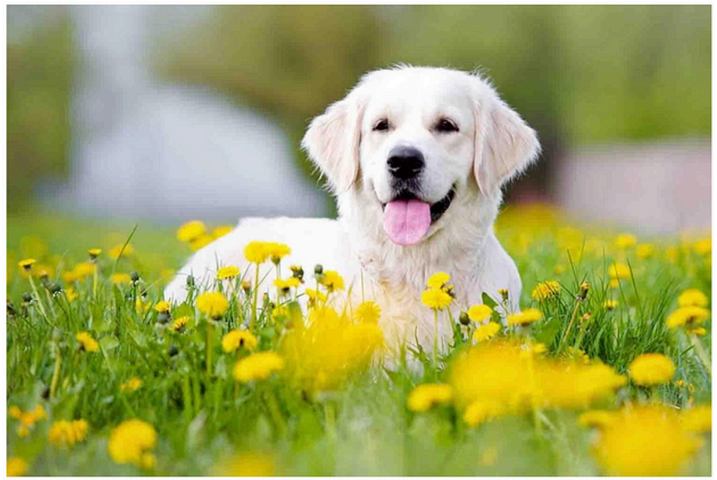


IMAGE SUPER-RESOLUTION

Image Super-Resolution aims to **increase the resolution** of an image while maintaining its content and details as much as possible. The end result is a **high-resolution version** of the original image. It's widely applied in areas like **medical imaging** and **satellite imaging**.



Low Resolution



High Resolution

IMAGE SEGMENTATION

Encoder-decoder architectures are extensively used in **image segmentation**, where the task is to **classify each pixel** of an image into predefined categories.

Image segmentation offers **more information** about an image than object detection, which draws a bounding box around the detected object, or image classification, which assigns a label to the object.

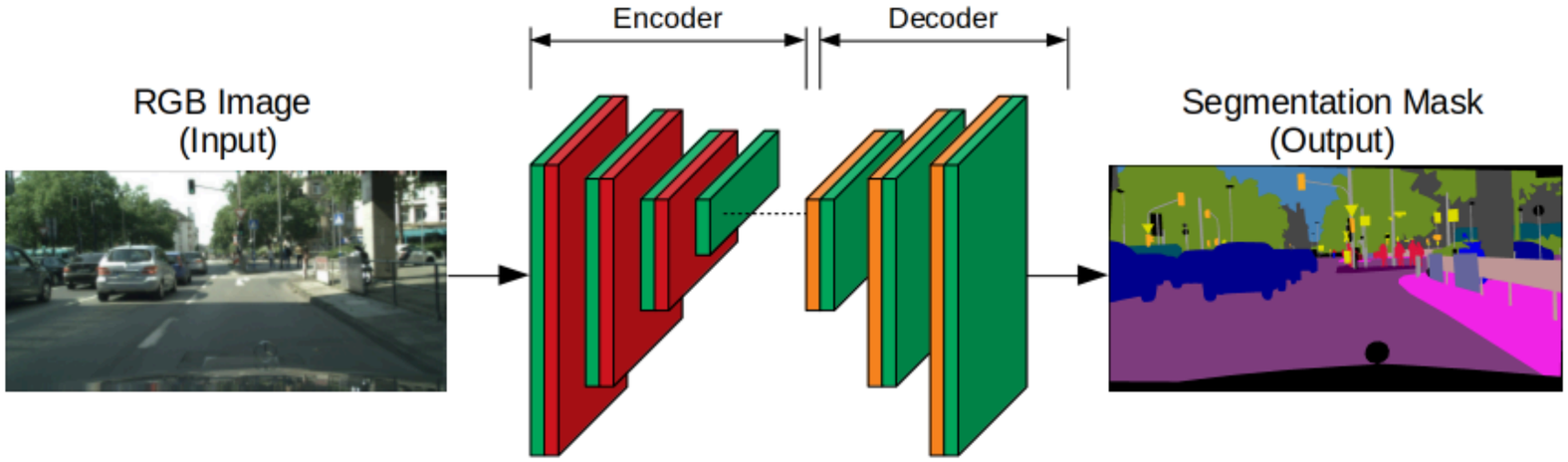
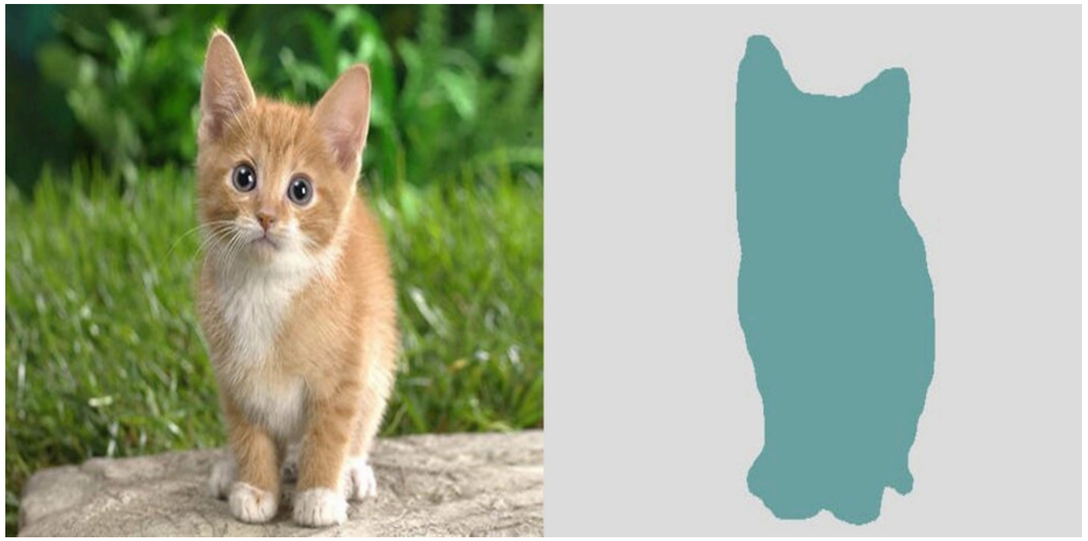


IMAGE SEGMENTATION

Image segmentation involves partitioning the image into smaller parts called **segments**. Segmentation involves understanding what is given in an image at a pixel level.

It provides **fine-grained information** about the image as well as the shapes and boundaries of the objects. The output of image segmentation is a mask where each element indicates which class that pixel belongs to.



U-NET

U-Net is a convolutional neural network architecture originally developed for biomedical image segmentation and introduced in 2015. The architecture has since gained widespread use in various segmentation tasks due to its simplicity, efficiency, and accuracy.

U-Net has a symmetrical structure with two main parts:

- **Encoder (Contracting Path)**: Extracts features by progressively reducing spatial dimensions while increasing the depth of feature maps.
- **Decoder (Expanding Path)**: Reconstructs spatial dimensions using upsampling operations and combines low-level spatial information with high-level features.

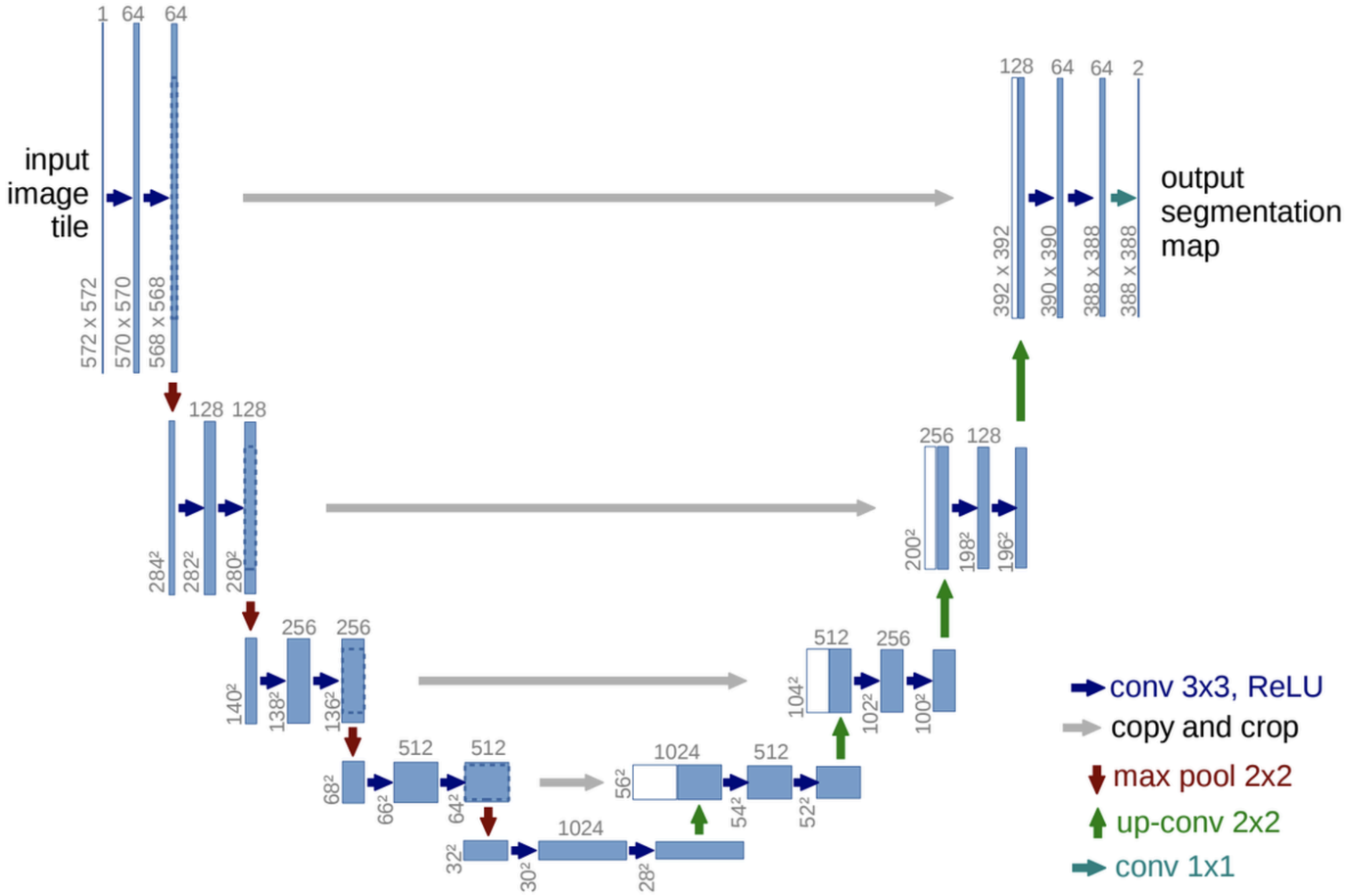
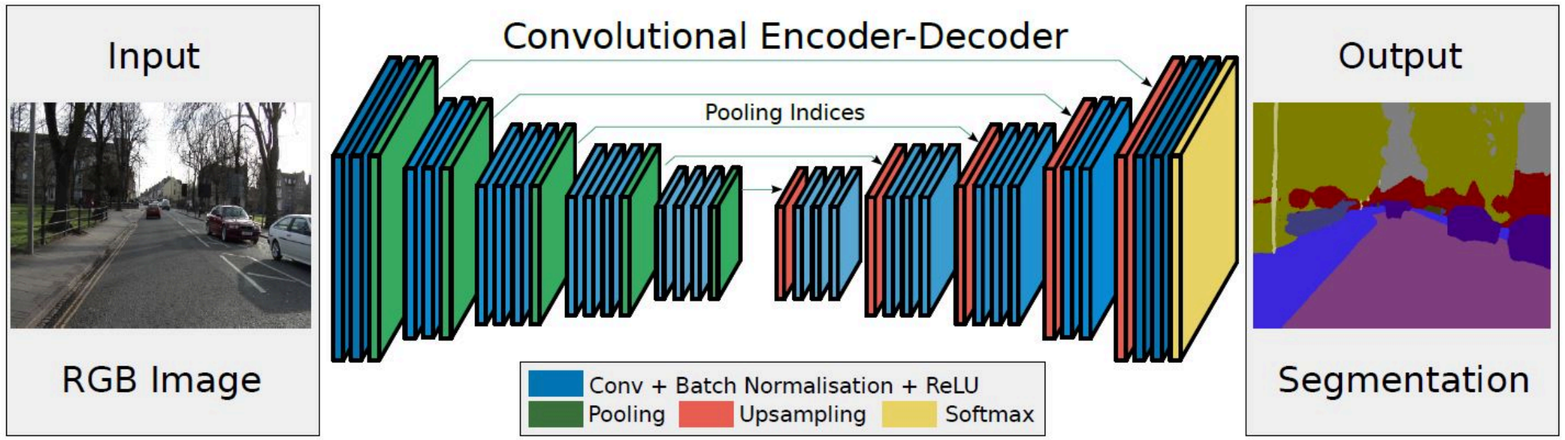


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

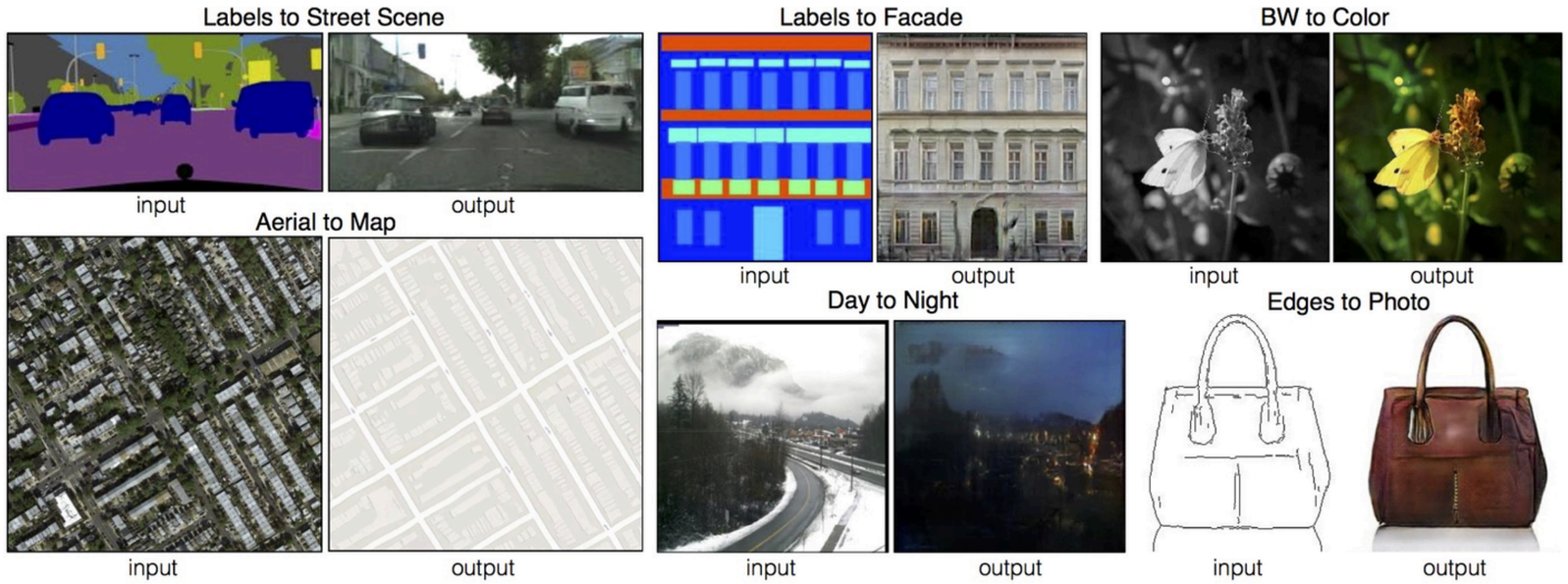
OTHER NOTABLE NETWORKS

- **SegNet**: a similar approach to U-net, is more efficient in terms of memory usage, but yields lower precision. It performs well in real-time applications where speed is crucial (e.g.: road segmentation for autonomous cars)



OTHER NOTABLE NETWORKS

- **Pix2Pix**: uses a U-Net-like architecture for its generator. The generator takes an input image, processes it through an encoder-decoder network, and outputs a transformed image. This transformation could be anything from turning a black-and-white image into color, to converting a sketch into a photorealistic image.



THANK YOU