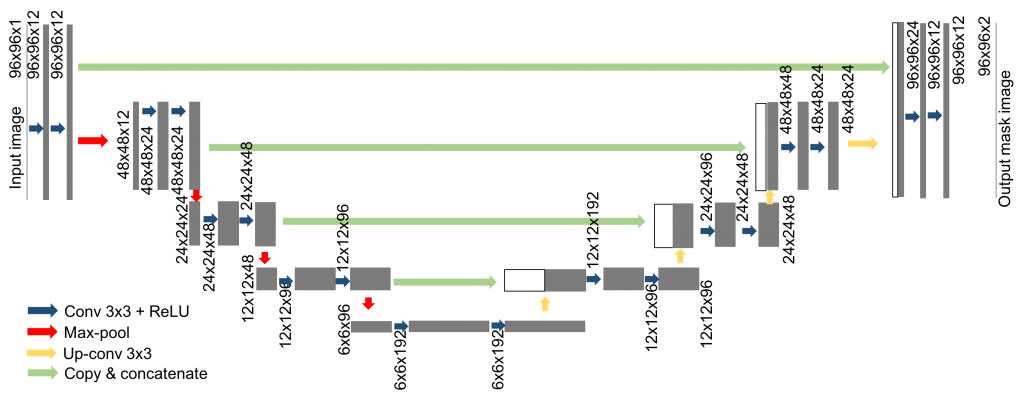
Crack detection using U-Net architecture

U-Net is a convolutional neural network based on the fully convolutional network and its architecture was modified and extended to work with fewer training images and to yield more precise segmentations. Segmentation of a 512x512 image should take less than a second on a modern GPU.

The U-Net Model:



* The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3×3 convolutions, each followed by a batchnormalization layer and a rectified linear unit (ReLU) activation and dropout and a 2×2 max pooling operation with stride 2 for downsampling. At each downsampling step we double the number of feature channels. The purpose of this contracting path is to capture the context of the input image in order to be able to do segmentation.
* Every step in the expansive path consists of an upsampling of the feature map followed by a 2×2 convolution (“up-convolution”) that halves the number of feature channels, a concatenation with the correspondingly feature map from the contracting path, and two 3×3 convolutions, each followed by batchnorm, dropout and a ReLU. The purpose of this expanding path is to enable precise localization combined with contextual information from the contracting path.
* At the final layer a 1×1 convolution is used to map each 16- component feature vector to the desired number of classes.

Relevant articles:

* Jenkins, Mark David, et al. "A Deep Convolutional Neural Network for Semantic Pixel-Wise Segmentation of Road and Pavement Surface Cracks." 2018 26th European Signal Processing Conference (EUSIPCO). IEEE, 2018.

(<https://www.eurasip.org/Proceedings/Eusipco/Eusipco2018/papers/1570437180.pdf>)

Model used: U-Net architecture with a final softmax layer that trains patches of the images from the original set in order to eliminate the useless background information before training.

Results: Image patches of size 48x48 are utilized and a total of 2000 random patches are extracted from each of the training images. The algorithm is evaluated on the CrackForest Dataset made of images of size 480x320. The network is trained on 100 of these images split into 80 training and 20 validation images. The implementation was carried out using Keras and Tensorflow and training was carried out over 100 epochs with a batch size of 34. Training takes approximately 3 hours and inference on the pre-processed test set of 18 images takes approximately 3 seconds. Precision: 92.46%

* Mosinska, Agata, et al. "Beyond the pixel-wise loss for topology-aware delineation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

(<http://openaccess.thecvf.com/content_cvpr_2018/papers/Mosinska_Beyond_the_Pixel-Wise_CVPR_2018_paper.pdf>)

Model used: A combination of the U-Net model and the VGG19 network in order to better determine possible errors that may appear on feature detection on the image.

Results: Images of cracks in road re split into 104 for training and 20 test images. The multiple shadows and cluttered background make their detection a challenging task. Patches of 450x450 pixels are used for training. The data is augmented mirroring and rotating the training images by 90◦, 180◦ and 270◦. Batch normalization is used for faster convergence and the Adam optimizer. Precision: 94.3%

* Escalona, Uriel, et al. "Fully Convolutional Networks for Automatic Pavement Crack Segmentation." Computación y Sistemas 23.2 (2019): 451-460.

(<https://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/viewFile/3047/2625>)

Model used: Three U-Net models, with 23, 11 and 7 convolutional layers in order to determine which one gives better results.

Results: Using the CFD dataset, made of images of size 480x320x3 (each having 3 different illuminations/shades) and the AigleRN dataset, made of 991x462 grey scale images, 100 images were used for training and 18 for testing. Precisions: 93.45%, 94.28%, 82.42%

* Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.

([https://arxiv.org/pdf/1505.04597.pdf)%e5%92%8c%5bTiramisu%5d(https://arxiv.org/abs/1611.09326.pdf](https://arxiv.org/pdf/1505.04597.pdf)%e5%92%8c%5bTiramisu%5d(https:/arxiv.org/abs/1611.09326.pdf))

Model used: U-Net on biomedical image segmentation

Results: Using 23 convolutional layers and augumented data by generating alternative images with deformations from a set of 30 512x512 images for training. Precision: 92.03%

* <https://sergioskar.github.io/Semantic_Segmentation/>

Relevant utilities:

* <https://data.mendeley.com/datasets/c7cpnw32j6/1>
* <https://github.com/zhixuhao/unet>
* <https://pypi.org/project/keras-unet/>
* <https://www.depends-on-the-definition.com/unet-keras-segmenting-images/>