Demystifying Attention U-net

30.06.2020

Recap: Soft Attention-Additive Attention

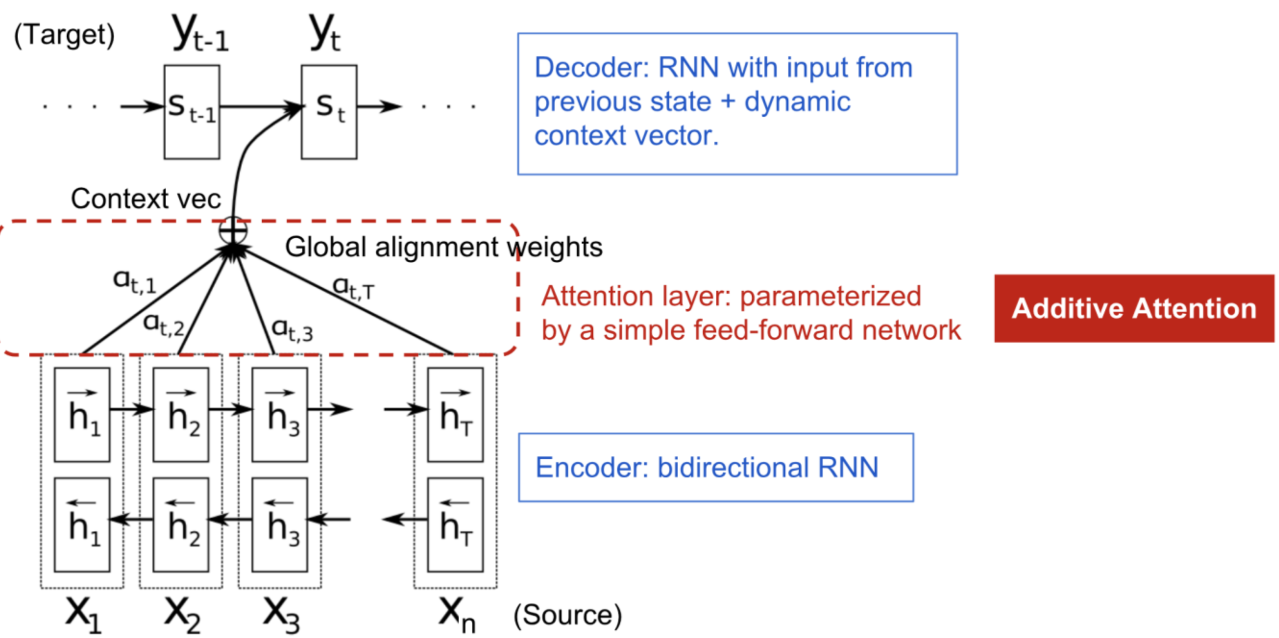


Figure 01. Diagram derived from Bahdanau et al., 2015 with addition information from [1]

$$\mathbf{c}_t = \sum_{i=1}^{n} \alpha_{t,i} \boldsymbol{h}_i \qquad \text{; Context vector for output } y_t$$

$$\alpha_{t,i} = \operatorname{align}(y_t, x_i) \qquad \text{; How well two words } y_t \text{ and } x_i \text{ are aligned.}$$

$$= \frac{\exp(\operatorname{score}(\boldsymbol{s}_{t-1}, \boldsymbol{h}_i))}{\sum_{i'=1}^{n} \exp(\operatorname{score}(\boldsymbol{s}_{t-1}, \boldsymbol{h}_{i'}))} \qquad \text{; Softmax of some predefined alignment score..}$$

$$\operatorname{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \mathbf{v}_a^\top \tanh(\mathbf{W}_a[\boldsymbol{s}_t; \boldsymbol{h}_i])$$

[1] Image source: https://lilianweng.github.io/lil-log/2018/06/24/attention-attention.html

Why U-Net?:

- Convolutional Neural Network (CNN) learn the feature mapping of an image and exploit it's salient features as the network depth increases
- Image is converted into vector for classification task – Works well in classification!
- Image segmentation -> Convert feature maps into vector but also reconstruct image from such vector!

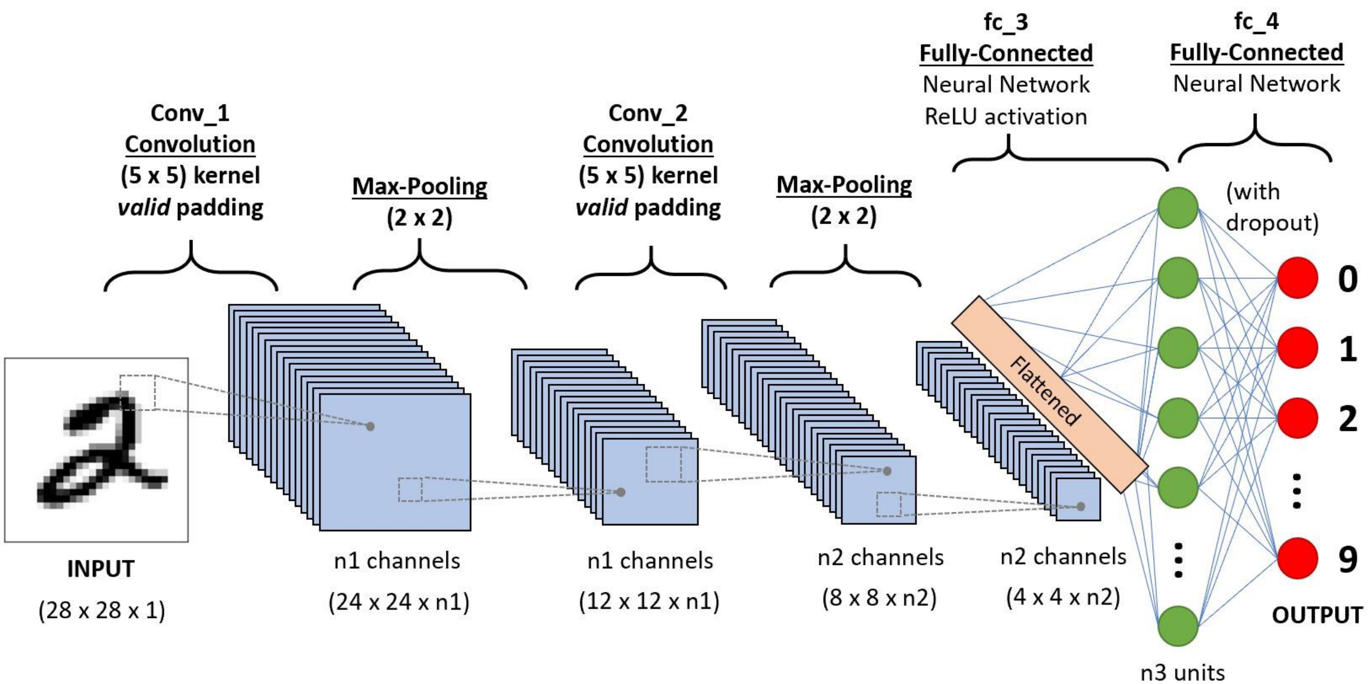


Figure 02. Classification task with Convolutional Neural Network [1]

[1] Image source: Ronneberger, O., Fischcher, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234-241. Springer(2017)

U-Net

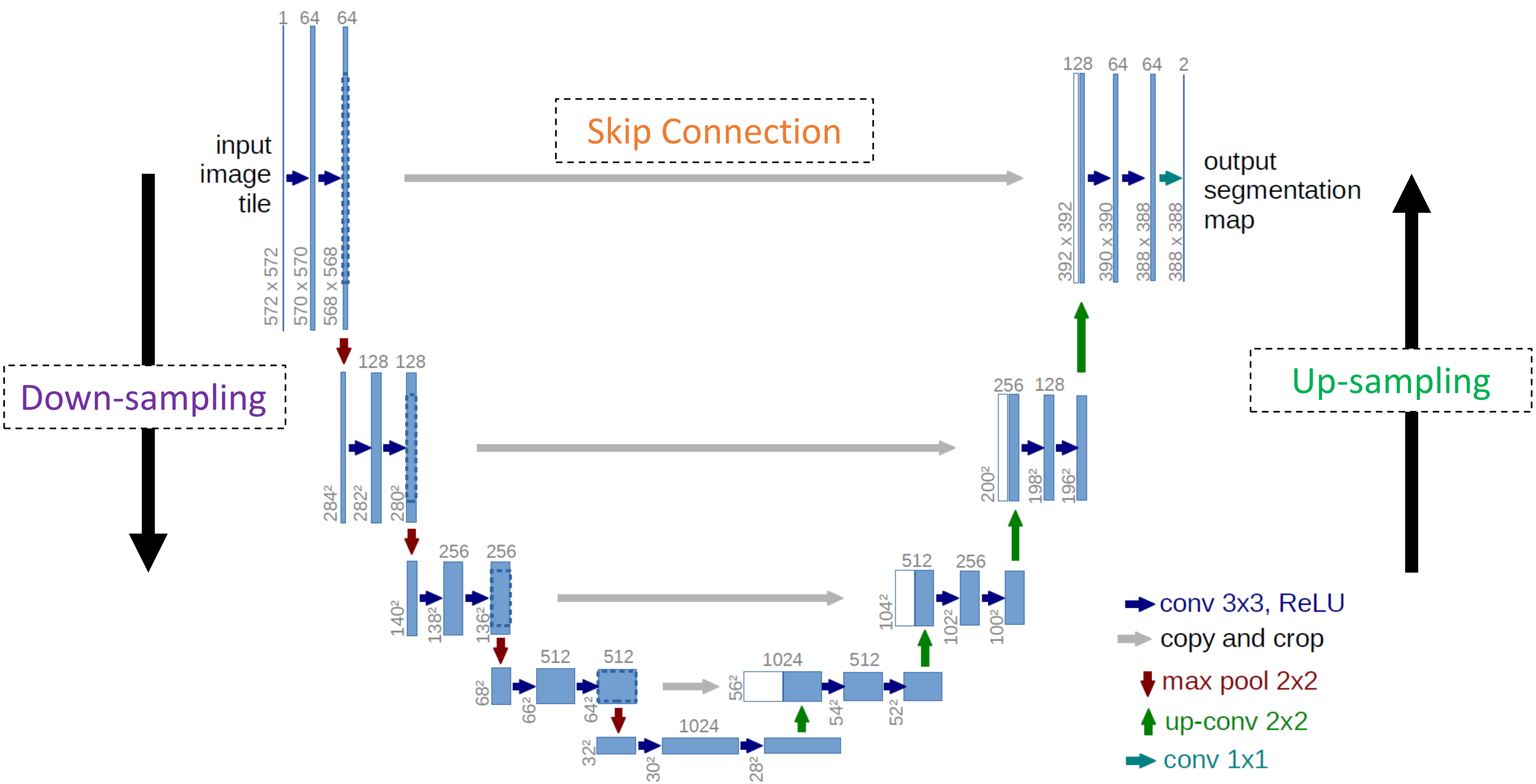


Figure 03. U-net model architecture derived from [1]

[1] Image source: Ronneberger, O., Fishcher, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234-241. Springer(2017)

Intuition for U-Net

- Down-Sampling:

  - Encode the input image into feature representations at multiple different levels

  - Learn local features (Localization)

- Up-Sampling:

  - Project discriminative features(lower resolution) learnt by the encoder onto the pixel space (higher resolution) for dense classification

  - Restore condensed feature maps to the original size of the input image by expanding the feature dimensions

  - Learn global features

- Skip Connections:

  - Concatenate higher resolution feature maps from earlier stage for better representations learning

  - Up-sampling is a sparse operation, need good prior from earlier stages to better represent localization

Drawbacks

- Rely on multi-stage cascaded CNNs when the target organs show variation in terms of shape and size
- Leads to excessive and redundant use of computational resources as well as model parameters
- Similar low-level features are repeatedly extracted by all models within the cascade
- Difficult to reduce false-positive predictions for small objects that show large shape variability

Attention U-Net

- Introducing Attention Gate(AG) to attend on salient features useful for specific task

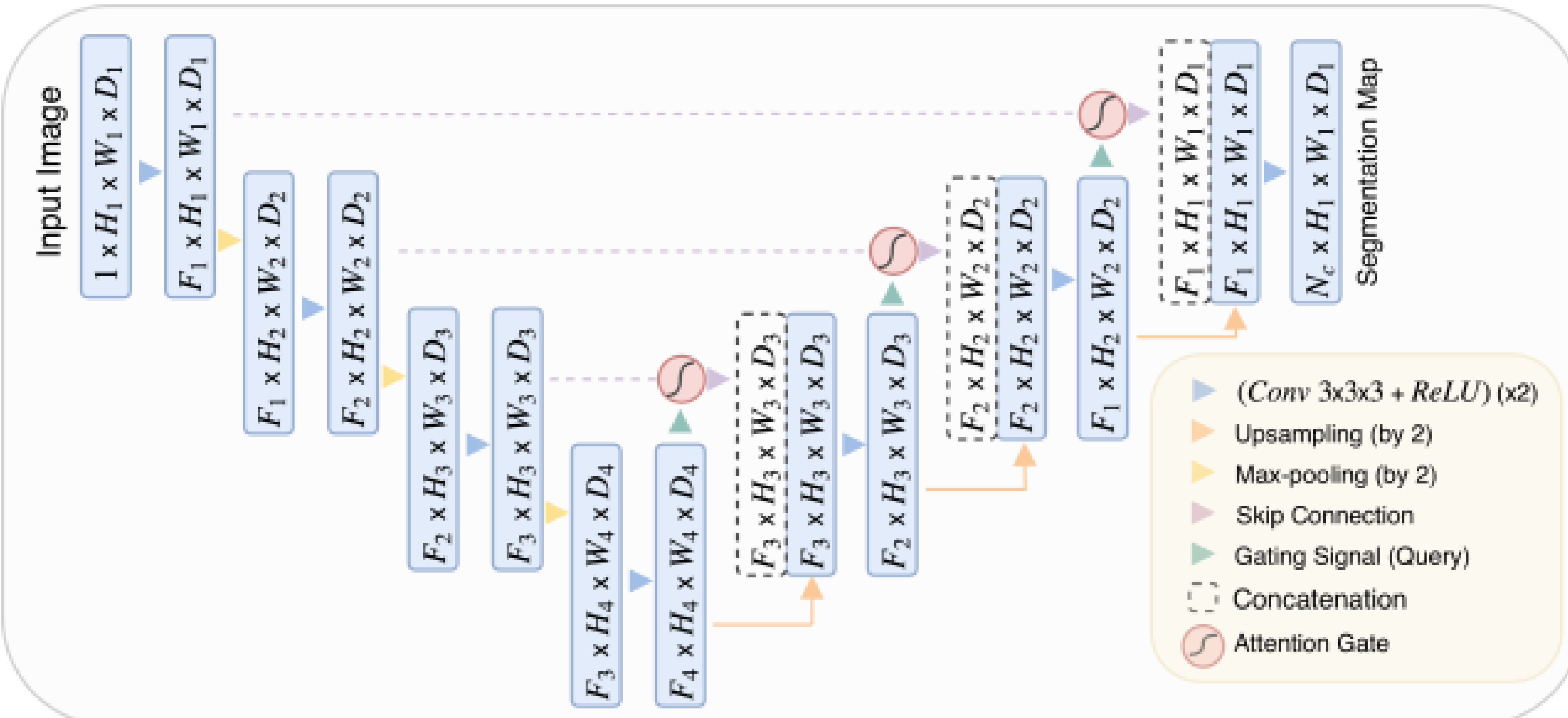
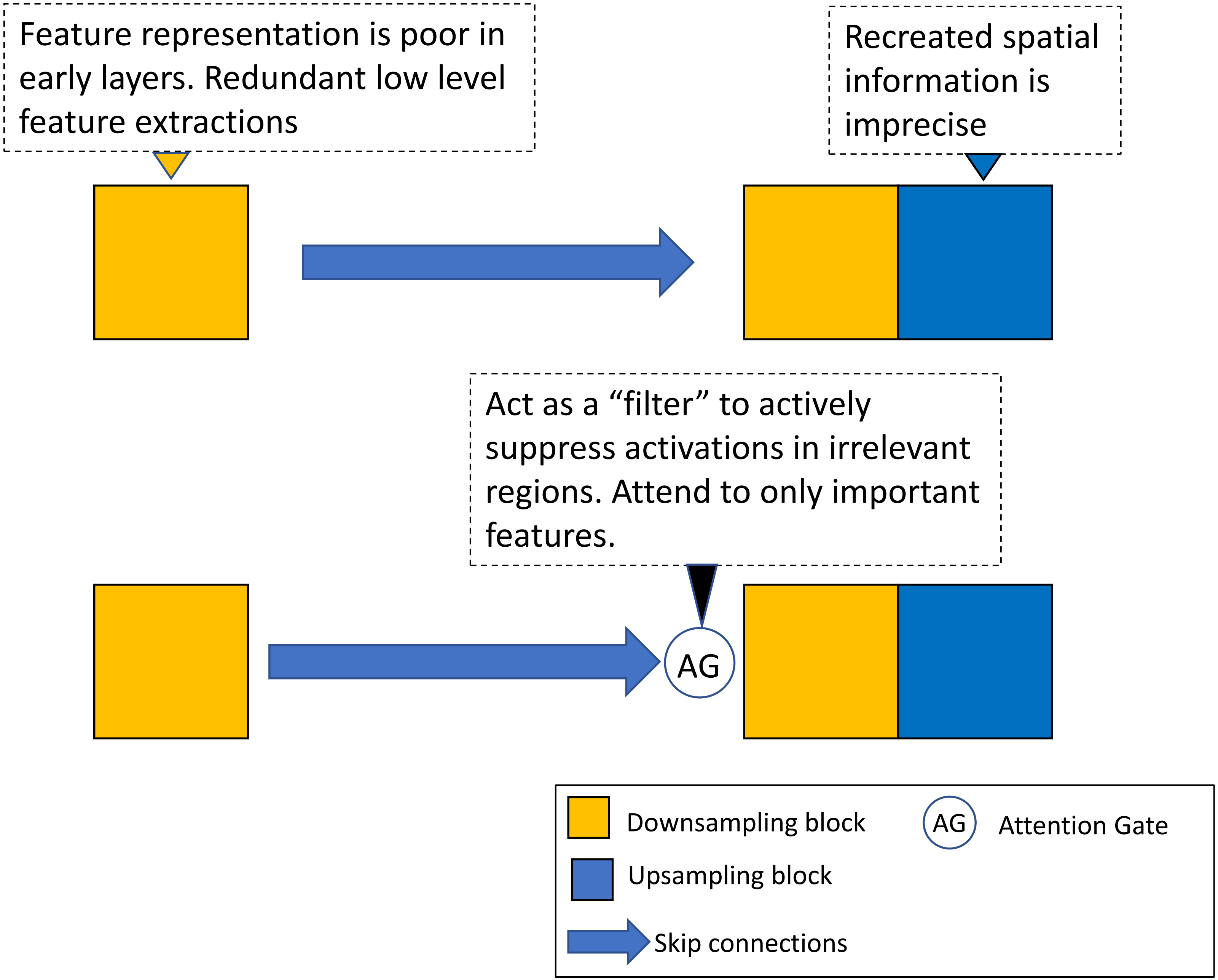


Figure 04. Attention U-net model architecture derived from [1]

[1] Image source: Oktay. et al.: Attention U-Net: Learning Where to Look for the Pancreas, 2018

Intuition for Attention Gates(AG)



Additive Attention Gate

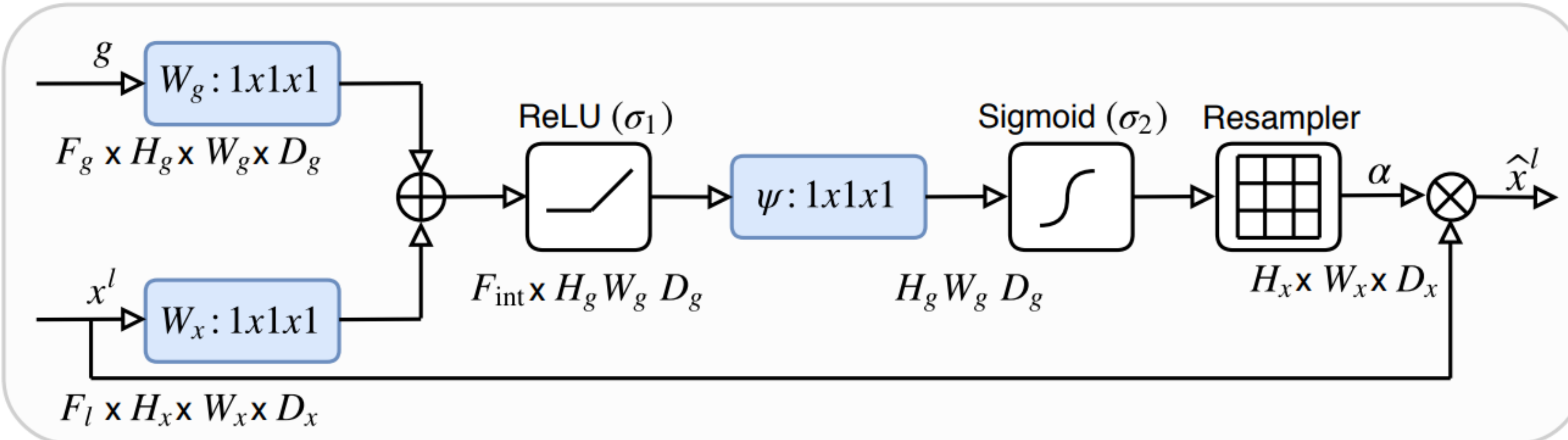


Figure 05. Attention gate schematic drawing from [1]

g: Gating signal (Next lowest layer/ Upsampling layer)

x^l: Input signal (Features from skip connection)

ψ: Linear transformation(Convolutional Layer)

$\oplus$: Additive operation

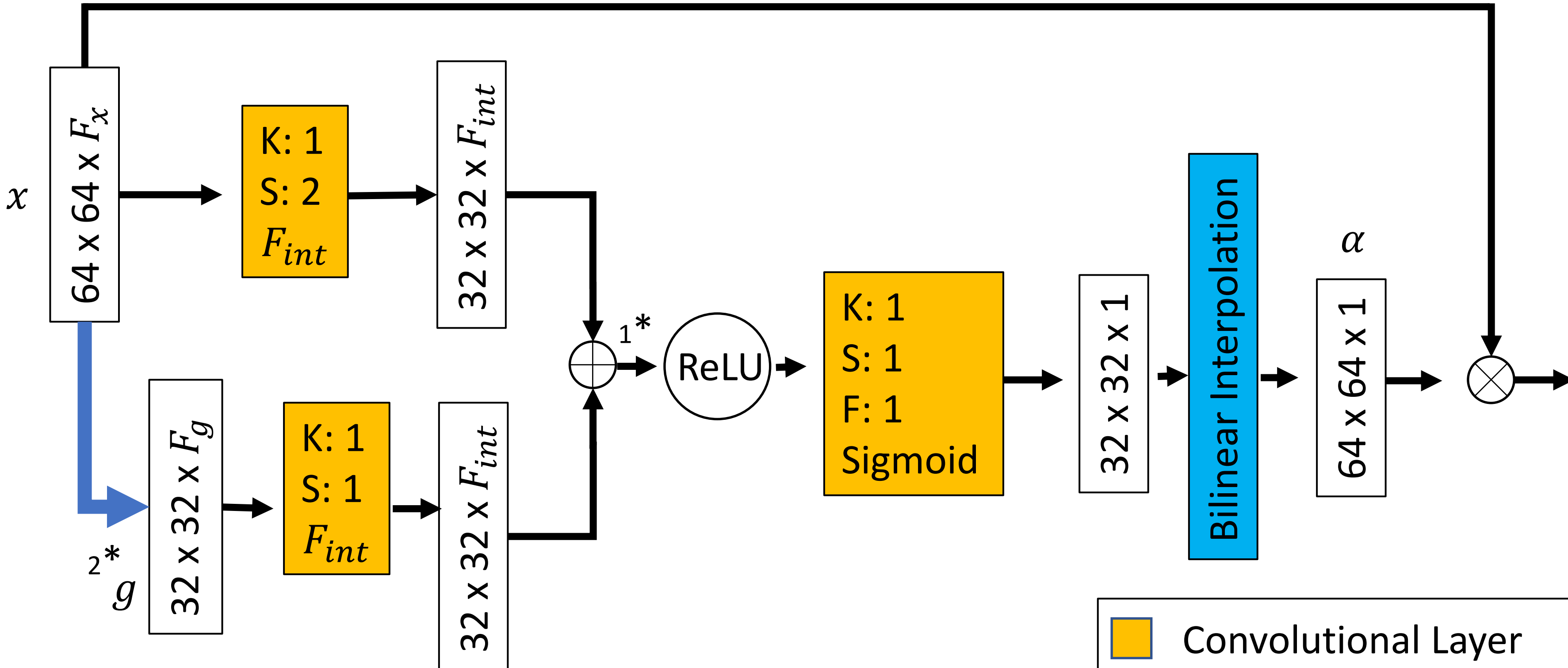
$\otimes$: Element wise multiplication operation

α: Attention coefficient $\in [0,1]$

Resampler: Trilinear interpolation (Upsampling)

[1] Image source: Oktay. et al.: Attention U-Net: Learning Where to Look for the Pancreas, 2018

Attention Gate Example (2D)



1*Element Wise Sum: Creates "superposition" affect where, aligned weights becoming larger while unaligned weights become relatively smaller

2*Grid Based Attention: Vector *g* from the upsampling path rather than the downsampling path, vector would have been conditioned to spatial information from multiple scales by previous attention gates

Convolutional Layer

Upsampling Layer

Downsampling Layer

[1] Reference: https://towardsdatascience.com/a-detailed-explanation-of-the-attention-u-net-b371a5590831

Summary

- AGs was designed to improve model sensitivity and accuracy by suppressing feature activations in irrelevant regions
- AGs generate soft region proposals implicitly and highlight salient features useful for specific task
- Enable access to introspection for U-Net through the visualization of attention weights

References

- Ronneberger, O., Fishcher, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234-241. Springer(2017)
- Oktay. et al.: Attention U-Net: Learning Where to Look for the Pancreas, 2018
- Robin Vinod: A detailed explanation of the Attention U-Net, https://towardsdatascience.com/a-detailed-explanation-of-the-attention-u-net-b371a5590831
- How U-net works?: https://developers.arcgis.com/python/guide/how-unet-works/
- Heet Sankesara, UNet: https://towardsdatascience.com/u-net-b229b32b4a71

Thank You for Your Kind Attention!