# Attention U-Net: Learning Where to Look for the Pancreas

O. Oktay[1,2], J. Schlemper[1], L. Le Folgoc[1], M. Lee[1,2], M. Heinrich[4], K. Misawa[3], K. Mori[3], S. McDonagh[1], N. Y Hammerla, B. Kainz[1], B. Glocker[1], and D. Rueckert[1]

[1] BioMedIA Group, Imperial College London, UK [2] HeartFlow Inc, California, USA
[3] Nagoya University, Aichi Cancer Center, JP [4] Institute of Medical Informatics, Uni of Lübeck, DE

## Contributions

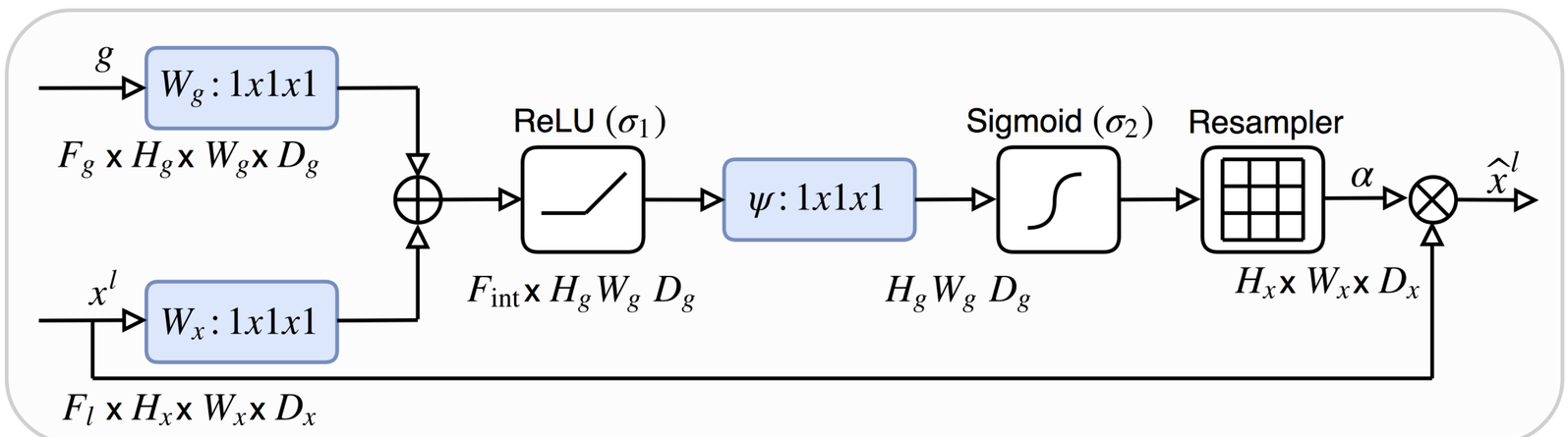
A novel soft-attention gate (AG) model is proposed for medical imaging that learns to focus on target structures of varying shapes and sizes without additional supervision. Models trained with AGs implicitly learn to suppress irrelevant regions in an image while highlighting salient features useful for a specific task. This enables:

⇒ The necessity of using an external organ localisation module as in cascaded networks can be eliminated whilst maintaining high prediction accuracy.

⇒ Improved model sensitivity via soft region proposals generated implicitly on-the-fly.

⇒ Better model interpretability, AGs highlight salient features useful for a specific task.

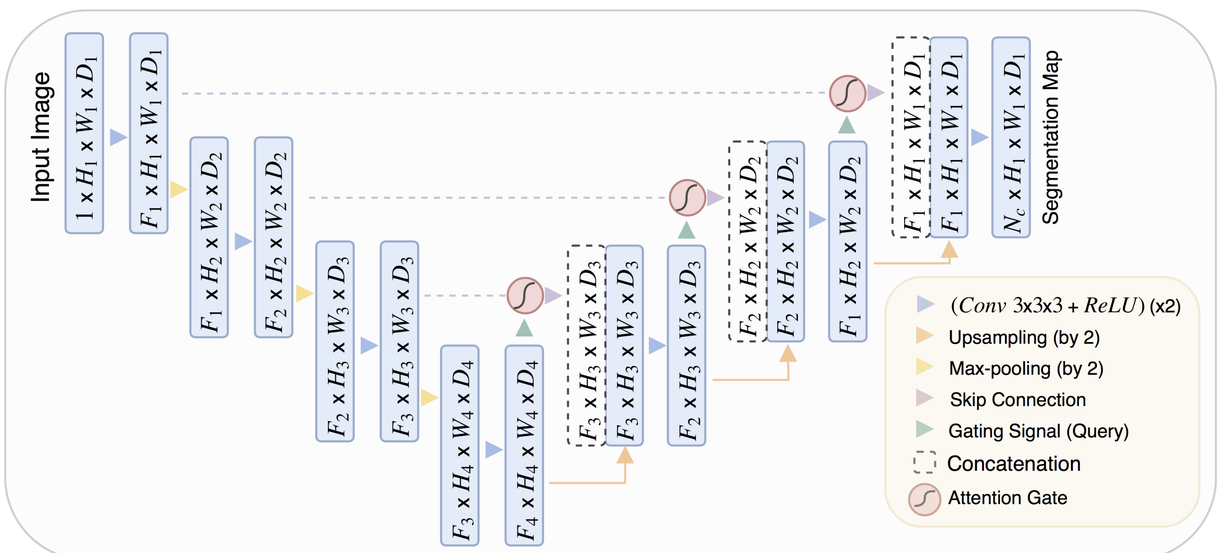
## Image-Grid Based Soft-Attention Gates

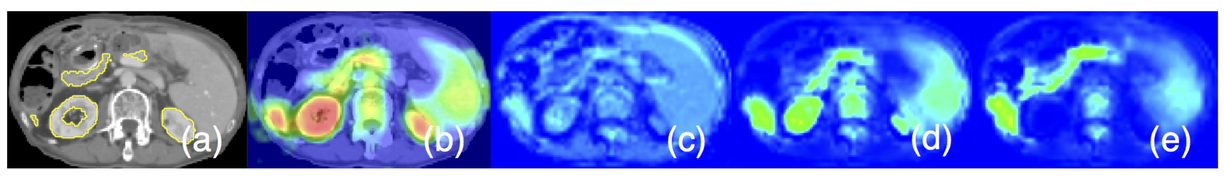



Input features ($x^l$) are scaled with per pixel attention coefficients ($\alpha \in [0, 1]$) computed in the attention gate. Image regions are selected by analysing both the activations ($x^l$) and contextual information provided by the gating signal ($g$) which is collected from a coarser scale. Linear transformations ($W_x$, $W_g$, $\psi$) are implemented as 3D channel-wise convolutions without any spatial support.

$$\alpha = \sigma_2(\psi^T(\sigma_1(W_x^T x + W_g^T g + b_g)) + b_\psi)$$

## Attention Gates In A CNN Segmentation Model




From left to right: Input image (a), attention-map (b), feature-map before the attention gate (c), filtered feature maps after the attention gates (d,e).

⇒ Contextual information ($g$) from coarser scales disambiguates local feature responses (skip connections) and highlights target organs/tissues.

⇒ At each image scale, the spatial search space is reduced and more localised towards target organs/tissues via progressive attention gating.

## Multi-Class CT Abdominal Segmentation (CT-150 and TCIA Pancreas Benchmarks)

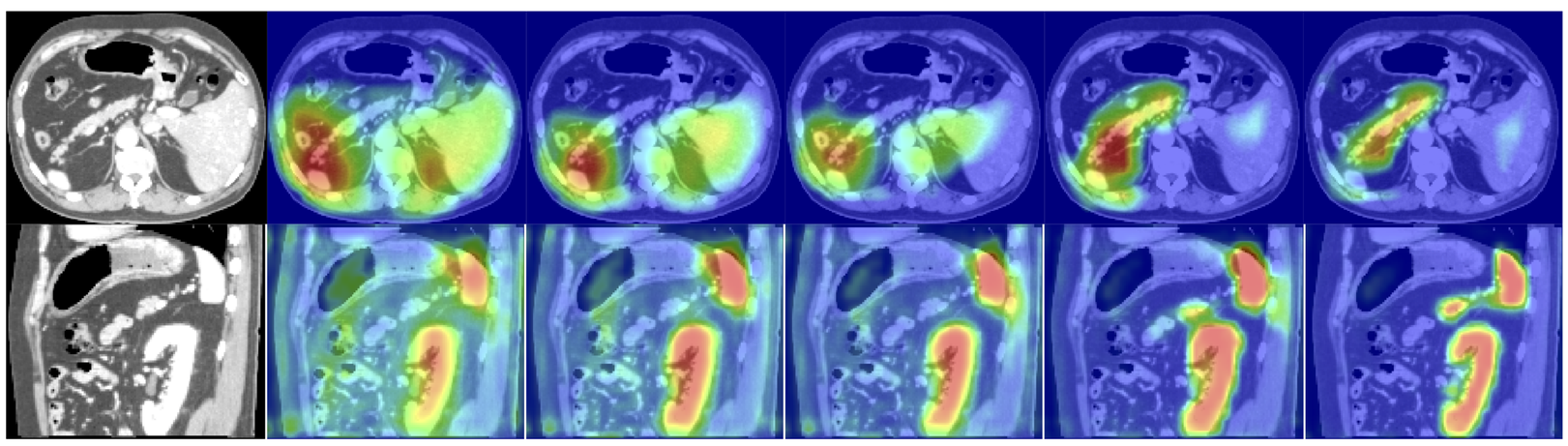
| Method (Train/Test Split) | U-Net (120/30) | Att U-Net (120/30) | U-Net (30/120) | Att U-Net (30/120) |
|---|---|---|---|---|
| CT-150 Dice Score | 0.814±0.116 | **0.840±0.087** | 0.741±0.137 | **0.767±0.132** |
| CT-150 Precision | 0.848±0.110 | 0.849±0.098 | 0.789±0.176 | **0.794±0.150** |
| CT-150 Recall | 0.806±0.126 | **0.841±0.092** | 0.743±0.179 | **0.762±0.145** |
| CT-150 S2S Dist (mm) | 2.358±1.464 | **1.920±1.284** | 3.765±3.452 | 3.507±3.814 |
| Number of Params | 5.88 M | 6.40 M | 5.88 M | 6.40 M |
| Inference Time | 0.167 s | 0.179 s | 0.167 s | 0.179 s |

| Method (Train/Test Split) | U-Net (61/21) | Att U-Net (61/21) | U-Net Finetune | Att U-Net Finetune |
|---|---|---|---|---|
| TCIA-82 Dice Score | 0.815±0.068 | 0.821±0.057 | 0.820±0.043 | **0.831±0.038** |
| TCIA-82 Precision | 0.815±0.105 | 0.815±0.093 | 0.824±0.070 | 0.825±0.073 |
| TCIA-82 Recall | 0.826±0.062 | **0.835±0.057** | 0.828±0.064 | **0.840±0.053** |
| TCIA-82 S2S Dist (mm) | 2.576±1.180 | **2.333±0.856** | 2.464±0.529 | **2.305±0.568** |

⇒ Standard U-Net and the proposed attention models are benchmarked on two different datasets (*CT*-150, *TCIA*-82), where models are used to segment the pancreas in 3D-CT abdominal images.

⇒ Attention gates (AGs) improve model accuracy and sensitivity as can be seen in recall values. AGs do not introduce significant computational overhead and do not require a large number of model parameters.

## Atttenion Coefficients Across Training Epochs



The proposed attention gate learns to localise target organs and filter background feature responses at each training epoch (From left to right, epoch={3, 6, 10, 60, 150}). The gate bias parameters are initialised with a constant value to start with a uniform attention distribution across all pixels.

## References

[1] Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. ICLR (2015)

[2] Jetley, S., Lord, N.A., Lee, N., Torr, P.H.: Learn to pay attention. ICLR (2018)

[3] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: NIPS (2017)