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

CS300 : Midsem Project Evaluation

Sreejita Das

Department of Computer Science and Engineering
Indian Institute of Information Technology, Guwahati

3 March, 2025

Problem Statement

- **Challenge**: Accurate segmentation of anatomical structures, such as the pancreas, in medical imaging remains an obstacle due to significant inter-patient variability in shape and size, low tissue contrast in CT scans, and the dependency on computationally intensive multi-stage convolutional neural networks (CNNs) that inefficiently localize regions of interest.
- **Goal**: Develop an innovative, single-model architecture—Attention U-Net—that integrates attention gates to autonomously emphasize salient features and suppress irrelevant regions, achieving enhanced segmentation precision with reduced computational overhead.

Literature Survey

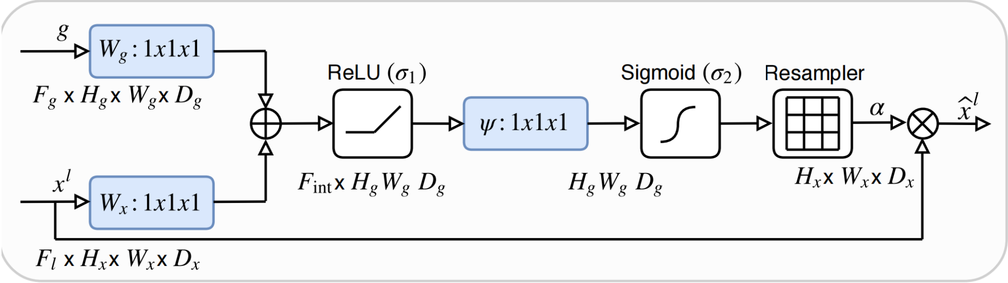
- **U-Net (Ronneberger, 2015):** Built a CNN for 2D medical image segmentation; basis for pancreas work in CT scans.
- **3D U-Net (Çiçek, 2016):** Adapted U-Net for 3D CT scans, connecting slices for full pancreas detail.
- **V-Net (Milletari, 2016):** Designed specifically for 3D volumetric data, for segmentation of small organs like the pancreas.
- **Dense Dilated Nets (Gibson, 2017):** Segments multiple organs in CT scans, utilizing dilated CNNs that capture the broader context for identifying multiple organs simultaneously.

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

- **Medical Image Segmentation:** Divides CT scans into regions, identifying organs for diagnosis and treatment.
- **Pancreas Challenges:** Low tissue contrast, high inter-patient variability and proximity to organs (e.g., spleen, stomach) complicate localization, often causing poor boundary detection and overlap confusion.
- **Attention U-Net:** Builds on U-Net (Ronneberger, 2015) by adding **Attention Gates (AGs)** in the skip connections. AGs weigh features from the encoder with context from the decoder, focusing on pancreas regions and suppressing noise in one model. Cascaded CNNs stack networks, increasing computational overhead and requiring extra object localization modules.
- **Motivation:** AGs enhance efficiency and precision, avoiding the resource-heavy complexity of cascaded CNNs for the tricky pancreas.

Model Architecture

Main Paper Link: click here



Methodology: Attention Mechanism

1. **Input Processing:** Two parallel inputs:
   - **Local features** x_i^l from finer layers
   - **Gating signal** g_i from coarser scales
2. **Feature Transformation:** Both inputs undergo linear transformations via 1x1x1 convolutions.
$$W_x^T x_i^l \quad \text{and} \quad W_g^T g_i$$
3. **Feature Combination and Non-Linearity:** Additively combined and passed through ReLU:
$$\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)$$
4. **Attention Coefficients:** Transformed via linear mapping, normalized by sigmoid:
$$\alpha_i^l = \sigma_2\left(\psi^T \sigma_1(W_x^T x_i^l + W_g^T g_i + b_g) + b_\psi\right)$$
5. **Gated Output:** Final gated output via element-wise multiplication:
$$\hat{x}_{i,c}^l = x_{i,c}^l \cdot \alpha_i^l$$

Training Results

- **Dataset:** Pancreas CT-82 (TCIA)
- **Patients:** 60 out of 82 total (till now)
- **Epochs:** 70 (till now)

Metric	Predicted Attention U-Net	Actual Attention U-Net
Dice Score	65%–70%	81.5% $\pm$ 6.8%
Precision	67%–71%	81.5% $\pm$ 9.3%
Recall	68%–72%	83.5% $\pm$ 5.7%
S2S Distance	3.0–3.5 mm	2.33 $\pm$ 0.86 mm

Future Improvement: Enhancing Attention U-Net

- **Limitation:** Single attention focus struggles with complex organ structures, leading to inaccurate segmentation of overlapping regions like the pancreas and spleen.
- **Proposed Upgrade:** Multi-Head Attention (MHA) assigns multiple attention heads to capture both fine-grained details and larger spatial relationships simultaneously. Each head processes the same input, but learns distinct attention weights, capturing diverse spatial relationships at once.

  Example: Head 1 zeroes in on the pancreas-stomach interface. Head 2 focuses on the pancreas tail and its subtle overlap with the spleen's blood vessels. Head 3 captures the broader context, like how the pancreas aligns with the duodenum, ensuring the big-picture shape stays coherent.
- **Benefit:** Tackles the pancreas's complexity head-on—handling overlaps, details, and context in parallel - where single attention falters with its one-at-a-time approach

Related Work: Hierarchical 3D Fully Convolutional Networks for Multi-Organ Segmentation

- **What They Did:** Used a 3D U-Net with a hierarchical two-step method, a **coarse-to-fine** approach to detect **seven abdominal organs** (artery, vein, liver, spleen, stomach, gallbladder, pancreas) in CT scans.
- **First Step:** Roughly outlines organs by scanning **40% of the image** with a **simple body mask** to focus on relevant regions, separating the background from the organs.
- **Second Step:** Zooms into **10% of the image** based on the first guess, sharpening edges for **precise boundaries**.
- **Key Wins:** Big accuracy boost for **small organs** and adapts to **new hospital testing data**.

Dice	liver	spleen	pancreas	liver	spleen	pancreas	liver	spleen	pancreas
Mean	93.6	89.7	68.5	94.9	91.4	81.2	95.4	92.8	82.2
Std	2.5	8.2	8.2	2.1	8.9	10.2	2.0	8.0	10.2
Median	94.2	91.8	70.3	95.4	94.2	83.1	96.0	95.4	84.5
Min	78.2	20.6	32.0	80.4	22.3	1.9	80.9	21.7	1.8
Max	96.8	95.7	82.3	97.3	97.4	91.3	97.7	98.1	92.2

Conclusion

This project examined Attention U-Net, a deep learning model from recent research that refines pancreas segmentation in CT scans by embedding attention gates into the classic U-Net framework. It tackles persistent issues like unclear boundaries and organ overlaps, offering a more precise alternative to traditional bulky multi-stage models. The approach stands out for its balance of accuracy and efficiency, signaling a promising step toward smarter AI-driven diagnostics. Looking ahead, future work could broaden its applicability across diverse medical datasets and integrate clinical perspectives to enhance its real-world value.

References

Gibson, E., Giganti, F., Hu, Y., et al. (2017).
Automatic multi-organ segmentation with dense dilated networks.
In *MICCAI*.

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016).
V-net: Fully convolutional neural networks for volumetric medical image segmentation.
In *3DV*.

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

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016).
3d u-net: Learning dense volumetric segmentation from sparse annotation.
MICCAI.

Thank You