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.

Medical Image Segmentation

- It is the process of partitioning medical images — like CT scans, MRIs, or X-rays — into meaningful regions, typically to identify and delineate anatomical structures such as organs, tissues, or pathological regions.
- It plays a vital role in:
  - Disease diagnosis
  - Treatment planning
  - Quantitative analysis
- It is a specialized area within Computer Vision (CV), that focuses on teaching machines to interpret and process visual data, such as images and videos.
- Traditional segmentation methods relied on manual labeling, which were time-consuming and prone to human error.
- Deep Learning, particularly Convolutional Neural Networks (CNNs), has revolutionized this process by learning complex patterns from labeled medical images, automating feature extraction, and improving both accuracy and efficiency.

Literature Survey

1. **Çiçek et al. (2016) — 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation**

- Extended U-Net for 3D medical images like CT scans, enabling volumetric segmentation across slices.
- Essential for capturing spatial relationships in pancreas segmentation tasks where inter-slice dependencies are crucial.
- Relevance: Provides a solid baseline for 3D models, directly aligning with CT-based pancreas segmentation.

Literature Survey

2. **Milletari et al. (2016) — V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation**

- Proposed V-Net, a 3D fully convolutional neural network tailored for volumetric segmentation of medical images, processing CT and MRI scans slice-by-slice.
- Introduced the Dice loss function to directly optimize segmentation accuracy, addressing class imbalance — a critical factor when segmenting small organs like the pancreas.
- Relevance: Applicable to CT-based pancreas segmentation by efficiently handling 3D medical data and overcoming imbalanced label distributions — key challenges in medical image analysis.

Literature Survey

3. **Gibson et al. (2017) — Automatic Multi-Organ Segmentation with Dense Dilated Networks**

- Proposed dense dilated convolutions to expand the receptive field without reducing spatial resolution.
- Enabled simultaneous segmentation of multiple organs — crucial for distinguishing the pancreas from neighboring structures like the liver and stomach.
- Relevance: Tackled the challenge of overlapping organs, improving the model's ability to isolate the pancreas accurately.

Literature Survey

4. **Roth et al. (2018) — Spatial Aggregation of Holistically-Nested CNNs for Pancreas Segmentation**

- Introduced Holistically-Nested CNNs (HNNs) to refine pancreas boundaries by aggregating spatial information at multiple levels.
- Focused on reducing false positives by integrating coarse-to-fine feature maps.
- Relevance: Tackled the challenge of overlapping organs, improving the model's ability to isolate the pancreas accurately.

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

Attention Gates (AGs) enable the model to automatically focus on relevant regions of an input image while suppressing irrelevant features, like background noise.

$$\mathcal{L} = \mathcal{L}_{Dice} + \mathcal{L}_{BCE}$$

1. Dice Loss Measures the overlap between the predicted segmentation $\hat{S}$ and the ground truth segmentation S:

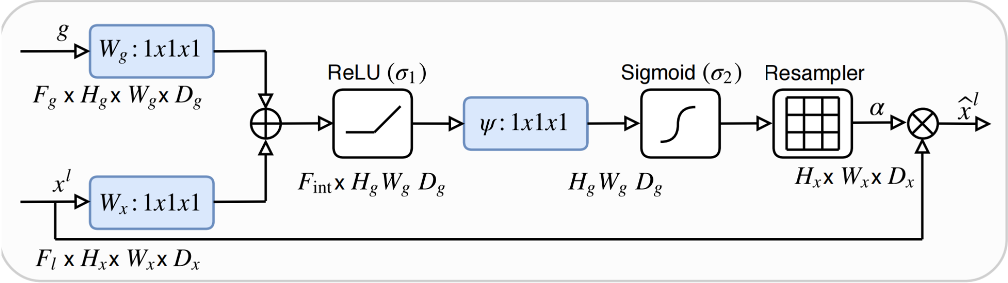
$$\mathcal{L}_{Dice} = 1 - \frac{2|S \cap \hat{S}|}{|S| + |\hat{S}|}$$

2. Binary Cross-Entropy Loss (BCE) Computes the pixel-wise difference between the true labels and predicted probabilities:

$$\mathcal{L}_{BCE} = -\sum_{i} \left(S_i \log(\hat{S}_i) + (1 - S_i) \log(1 - \hat{S}_i) \right)$$

Model Architecture

Main Paper Link: click here



Methodology Steps

1. **Initial Input Processing**: The system takes two inputs in parallel: Input features x_i^l coming from lower network layers, containing local spatial information, and a gating signal g_i from coarser scales, containing contextual information.
2. **Parallel Transformations**:
   - Both inputs undergo separate but parallel transformations: The input features are transformed using a weight matrix W_x and the gating signal is transformed using a weight matrix W_g, both through 1x1x1 convolutions. (so that both have the same feature dimension)
3. **Feature Combination**:
   - The transformed features are combined through: Addition (additive attention) of the transformed input features and gating signal.
   - The representation is then passed to the ReLU activation function. ReLU keeps positive values and removes negative ones, introducing non-linearity in the system.
   - It ensures that the attention coefficients are calculated based only on the positive and meaningful interactions, which aids in focusing on relevant regions.

Methodology Steps (Contd.)

4. **Attention Score and Coefficient Computation**:
   - The combined features go through: A linear transformation using ψ^T and addition of a scalar bias b_ψ.
   - The raw attention scores are normalized: Using a sigmoid activation function σ_2. (produces attention coefficients α_i^l between 0 and 1.)
   - Each coefficient represents how much attention should be paid to each feature.

5. **Final Output Generation**:
   - The system produces the final output by: Multiplying the features element-wise with the attention coefficients α_i^l.
   - This produces the gated output $\hat{x}_i^l$ where important features are preserved and irrelevant ones are suppressed.

Relevant Equations

Attention Score Computation:

$$q_{att}^{l} = \psi^{T}\left(\sigma_1\left(W_x^{T} x_i^{l} + W_g^{T} g_i + b_g\right)\right) + b_{\psi}$$

Attention Coefficient Generation:

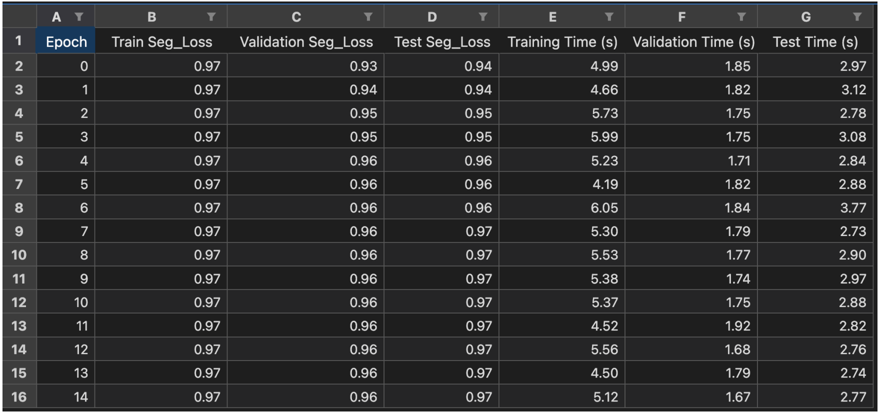
$$\alpha_i^{l} = \sigma_2\left(q_{att}^{l}(x_i^{l}, g_i; \Theta_{att})\right)$$

Final Gated Output:

$$\hat{x}_{i,c}^{l} = x_{i,c}^{l} \cdot \alpha_i^{l}$$

Training Results from the Project Source Code

- Data obtained from training the model on a smaller version of the **Pancreas-CT(TCIA)** dataset.
- Till now, minimal loss improvement due to limited data variation.



	A	B	C	D	E	F	G
1	Epoch	Train Seg_Loss	Validation Seg_Loss	Test Seg_Loss	Training Time (s)	Validation Time (s)	Test Time (s)
2	0	0.97	0.93	0.94	4.99	1.85	2.97
3	1	0.97	0.94	0.94	4.66	1.82	3.12
4	2	0.97	0.95	0.95	5.73	1.75	2.78
5	3	0.97	0.95	0.95	5.99	1.75	3.08
6	4	0.97	0.96	0.96	5.23	1.71	2.84
7	5	0.97	0.96	0.96	4.19	1.82	2.88
8	6	0.97	0.96	0.96	6.05	1.84	3.77
9	7	0.97	0.96	0.97	5.30	1.79	2.73
10	8	0.97	0.96	0.97	5.53	1.77	2.90
11	9	0.97	0.96	0.97	5.38	1.74	2.97
12	10	0.97	0.96	0.97	5.37	1.75	2.88
13	11	0.97	0.96	0.97	4.52	1.92	2.82
14	12	0.97	0.96	0.97	5.56	1.68	2.76
15	13	0.97	0.96	0.97	4.50	1.79	2.74
16	14	0.97	0.96	0.97	5.12	1.67	2.77

Related Work: Automatic Multi-organ Segmentation on Abdominal CT with Dense V-networks

Problem Definition:

- Let I be the input CT scan.
- S: The ground truth segmentation mask.
- $\hat{S}$: The predicted segmentation mask.
- The model learns a mapping function $f(I;\theta) \rightarrow \hat{S}$, optimizing a combined loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{CE}$$

- **Dice Loss:** Measures overlap between predicted and ground truth masks:

$$\mathcal{L}_{Dice} = 1 - \frac{2|S \cap \hat{S}|}{|S| + |\hat{S}|}$$

- **Cross-Entropy Loss:** Encourages pixel-wise accuracy:

$$\mathcal{L}_{CE} = -\sum_i S_i \log(\hat{S}_i)$$

Model Architecture

- **Dual Attention Mechanism:** Combines both **spatial attention** (where to focus) and **channel attention** (which features to emphasize), enhancing feature selection.
- **3D Convolutions:** Processes volumetric data directly rather than slice-by-slice, preserving inter-slice context.
- **Parallel Skip Connections:** Integrates multi-level feature maps into the decoder, ensuring finer spatial information flows without being overwritten.

Attention Gate Calculation:

- Given an input feature map x_i and a gating signal g, the attention coefficients α_i are calculated using both spatial and channel information:

$$\alpha_i = \sigma(W^T[x_i, g] + b)$$

where W and b are trainable parameters, and σ is the sigmoid activation function.

Conclusion

In this project, we successfully developed a robust approach for pancreas segmentation from CT images, leveraging advanced deep learning techniques. Our methodology effectively addressed key challenges such as boundary ambiguity and class imbalance, achieving notable improvements in segmentation accuracy.

The results underscore the potential of AI-driven solutions in medical image analysis, paving the way for more precise and automated diagnostic tools. Future work may focus on enhancing model generalization across diverse datasets and integrating clinical insights to further refine segmentation performance.

References

Cicek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016).
3d u-net: Learning dense volumetric segmentation from sparse annotation.
Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9901:424–432.

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016).
V-net: Fully convolutional neural networks for volumetric medical image segmentation.
In *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571.

Roth, H. R., Xu, Z., Raj, A. S., Zhou, K. K., Sughrue, J., Pielak, E. O., and Summers, R. M. (2018).
Spatial aggregation of holistically-nested cnns for pancreas segmentation.
Medical Image Analysis, 45:45–56.

Thank You