RIT

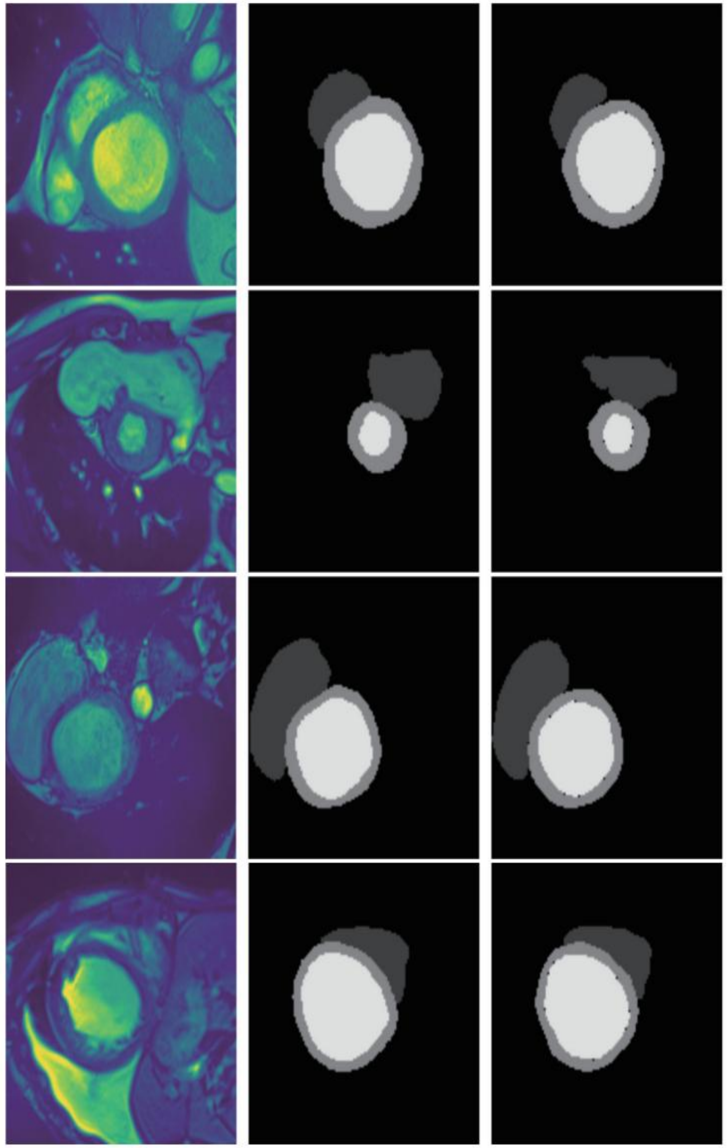
UNet and Variants for Medical Image Segmentation

Authors: Walid Ehab, Lina Huang, and Yongmin Li

Presentation by: Pratik Bhangale

Background

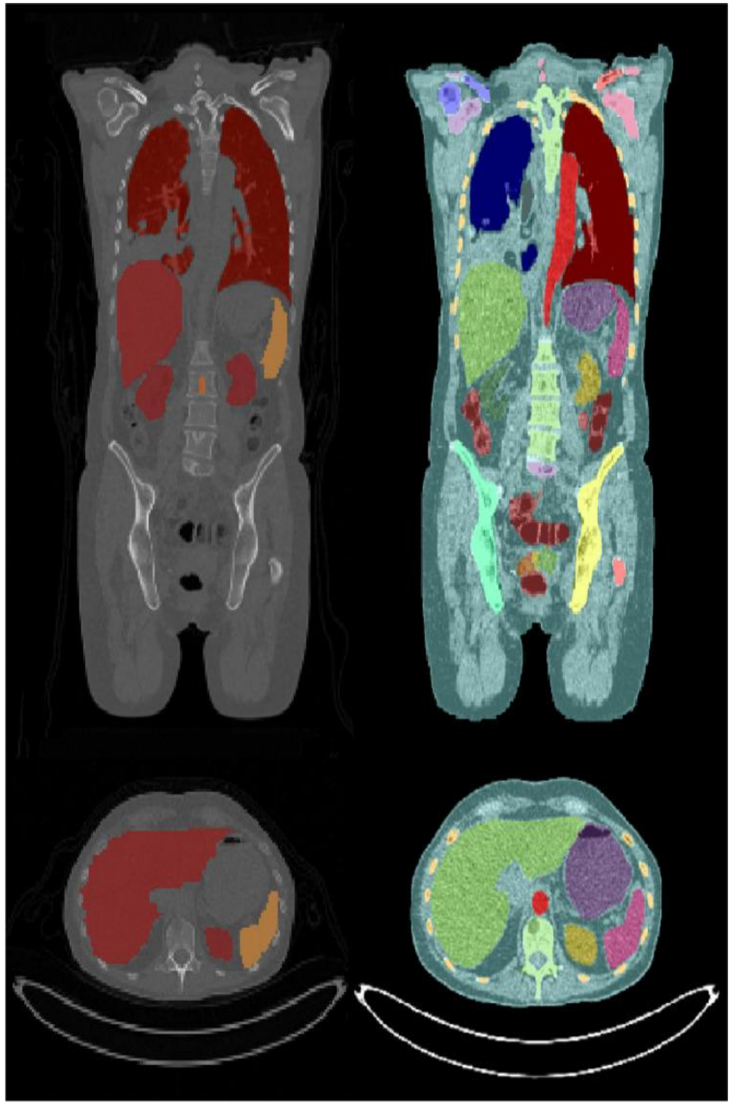
- **Medical Imaging & Segmentation**
  - Plays a critical role in modern healthcare (non-invasive visualization, early diagnosis)
  - Common challenges: organ/tumor shape variability, low contrasts, high noise.
- **Traditional Segmentation Approaches**
  - Thresholding, region-based, active contours, graph-cut, level-set methods
  - Often require heavy handcrafted features or precise parameter tuning
- **Rise of Deep Learning**
  - Convolutional Neural Networks (CNN) excel at automatically learning hierarchical features.
  - Fully Convolutional Networks (FCN), UNet, and attention-based models.
- **UNet as a Key Breakthrough**
  - U-shaped encoder-decoder network with skip connections.
  - Became a gold standard for biomedical/medical image segmentation tasks.



Brain tumor Segmentation

Motivation

- **Need for Automation**
  - Manual segmentation is time-consuming, error-prone, and subject to inter-observer variability.
- **Handling Complex Anatomies**
  - Different organs (e.g., heart, brain) or lesions (tumors, polyps) have varied sizes, shapes, and intensities.
- **Class Imbalance**
  - Large background regions vs. small lesion/organ areas; standard loss functions struggle in such cases.
- **Improving Clinical Workflows**
  - Consistency in segmentation can improve disease detection and patient outcomes.
- **Reducing Legal and Financial Risks**
  - Misdiagnoses, particularly in cancer detection, have led to hospitals facing lawsuits due to incorrect or delayed diagnoses, emphasizing the need for precise and automated medical image segmentation to minimize such errors and liabilities.
- **Empirical/Methodology Insights**
  - Deep learning models must address differing imaging modalities (MRI, CT, Ultrasound).
  - Careful architectural choices, loss design (focal loss, dice loss), and preprocessing are crucial.



Whole organ segmentation

Method

UNet

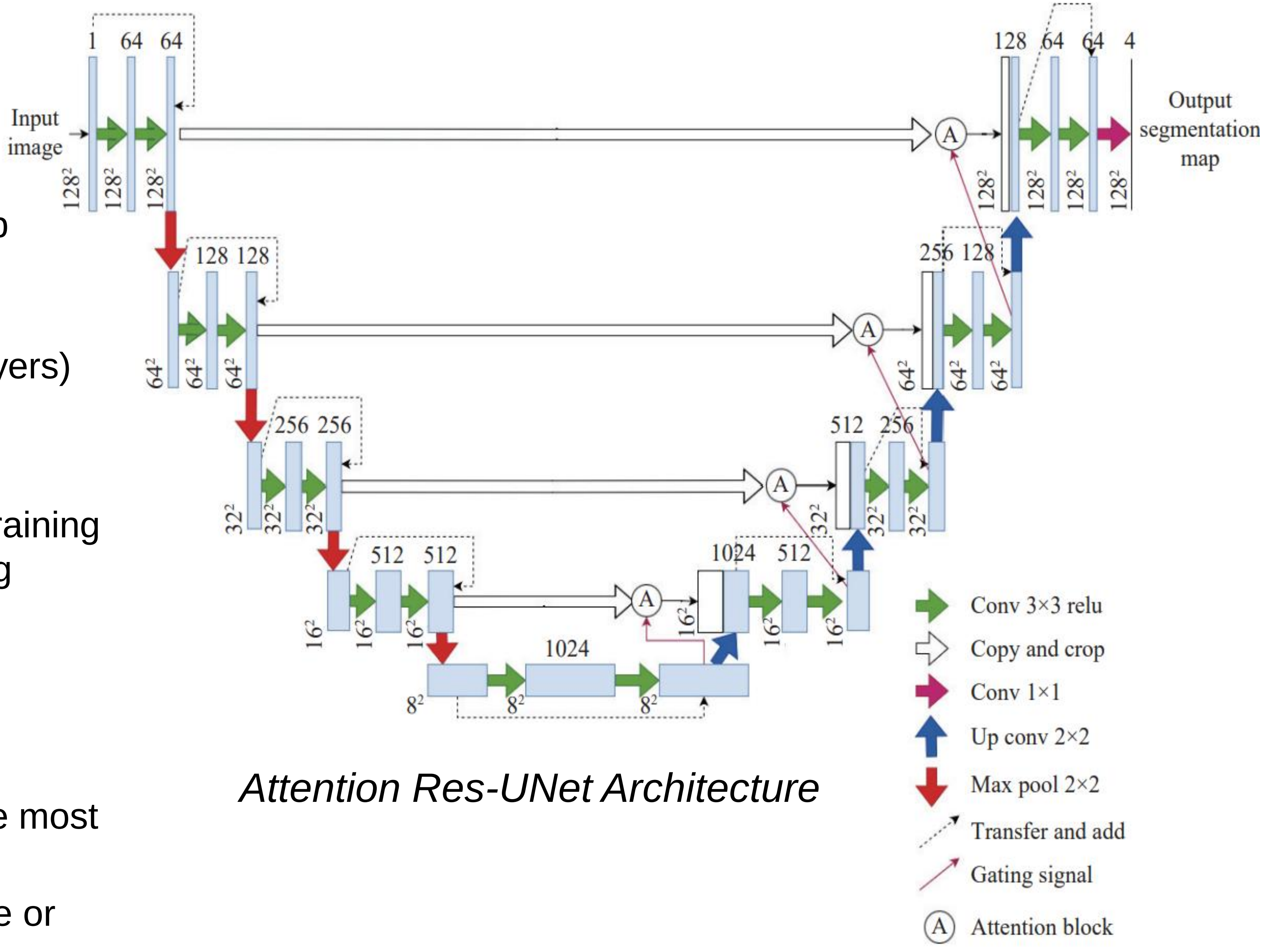
- Encoder-decoder structure with skip connections
- Captures both global context (deep layers) and local details (shallow layers)
- Highly effective in biomedical tasks

Res-Unet

- Integrates residual blocks to ease training deeper layers and combat vanishing gradients
- Residual skip connections → better feature reuse, stable convergence

Attention Res-Unet

- Adds attention gates to focus on the most relevant regions in feature maps
- Improves segmentation along subtle or small-scale boundaries



Attention Res-UNet Architecture

Technical Highlights

Loss Functions

- Binary/Categorical Focal Loss for imbalance
- Dice Coefficient used as a metric (and sometimes as a loss)

Training Details

- Optimizer: Adam with low learning rates (e.g., 1e-5)
- Regularization: Early stopping, reduced LR on plateau
- Preprocessing: Resizing, normalization, one-hot encoding for multi-class tasks

Challenges / Novelties

- Class imbalance handling → focal loss, data augmentation
- Fine-scale structures → skip connections + attention gating
- Resource constraints → flexible architecture design

The formula for the **Dice Coefficient** is:

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

Where:

- A and B are two sets.
- $|A|$ and $|B|$ are the sizes (cardinalities) of the sets.
- $|A \cap B|$ is the number of elements common to both sets.

The formula for **Binary Focal Loss** is:

$$FL(p_t) = -\alpha(1 - p_t)^{\gamma} \log(p_t)$$

Where:

- p_t is the predicted probability for the true class:

$$p_t = \begin{cases} p, & \text{if the true label } y = 1 \\ 1 - p, & \text{if the true label } y = 0 \end{cases}$$

- α is a weighting factor (optional) to balance class importance.
- γ (gamma) is the focusing parameter that reduces the loss for well-classified examples, typically set between 1 and 5.

Experiment Results (Binary)

Brain Tumor Segmentation

- Dataset: TCGA-LGG (FLAIR MRI, ~110 patients)
- Preprocessing: Resize (256×256), remove tumor-free scans, standardization
- Architectures: UNet, Res-UNet, Attention Res-UNet
- Loss Function: Binary Focal Loss (handling tumor vs. background imbalance)
- Training: Early stopping (~60–100 epochs)

Model	Dice Score	IoU	Recall
UNet	0.72	0.56	0.62
Res-UNet	**0.93**	**0.87**	0.94
Attention Res-UNet	0.92	0.86	**0.95**

Polyp Segmentation

- Dataset: CVC-ClinicDB (~612 images)
- Preprocessing: Resize (256×256), normalize
- Models & Training:
- Architectures: UNet, Res-UNet, Attention Res-UNet
- Loss Function: Binary Focal Loss, Adam optimizer

Model	Accuracy	Precision	Recall	Dice Score
UNet	96.8%	91.3%	73.3%	81.3%
Res-UNet	**97.1%**	**92.5%**	76.6%	**83.8%**
Attention Res-UNet	96.9%	88.1%	**78.8%**	83.2%

Experiment Results (Multi-Label)

Heart Segmentation (Multi-Label)

- Dataset: ACDC (~150 patient MRI volumes)
- Labels: Background (0), RV (1), Myocardium (2), LV (3)
- Preprocessing: Convert MRI "nifti" to slices, resize/crop (128×128), one-hot encoding

Architectures: UNet, Res-UNet, Attention Res-UNet

- Output Shape: (128×128×4), Softmax activation
- Loss Function: Categorical Focal Loss for class imbalance.

Model	Accuracy	Loss
UNet	**98.41%**	**1.00%**
Res-UNet	98.41%	1.09%
Attention Res-UNet	98.28%	1.44%

Dice Score for Heart Segmentation (Multi-Label)

Model	Background (0)	RV Cavity (1)	Myocardium (2)	LV Cavity (3)
UNet	**0.993**	0.906	**0.893**	**0.951**
Res-UNet	**0.993**	**0.920**	0.888	0.944
Attention Res-UNet	**0.993**	0.908	0.884	0.945

Conclusions

- U-Net, Res-U-Net, and Attention Res-U-Net all perform well, but Res-U-Net & Attention Res-U-Net handle fine details better.
- Focal loss reduces class imbalance, boosting recall and Dice scores for small or rare regions.
- U-Net converges faster but struggles with tiny targets; residual/attention blocks help preserve complex boundaries.
- For multi-class heart segmentation, U-Net & Res-U-Net excel on main classes, while Attention Res-U-Net captures underrepresented classes better.
- Cropping, normalization, and one-hot encoding are critical for reliable training and attention to subtle features.

Insights for Future Work:

- Investigate more complex or 3D architectures and datasets to better capture volumetric details.
- Explore data augmentation and advanced transfer learning to enhance model robustness.

Contributions to the Field:

- Provides a modern benchmark for comparing U-Net variants in medical segmentation tasks.
- Offers practical guidelines on preprocessing, loss functions, and architecture choices for researchers.

Thank You.

Any Questions?