Analysis of Competitor Methods for Thyroid Nodule Segmentation

# 1. Competitors Overview

## DPAM-UNet++

**Topic**: Dual-path Attention Mechanism Enhanced UNet++ for Thyroid Nodule Segmentation

**Why it's a direct competitor?** Uses U-Net++ with attention specifically for thyroid segmentation, aligning closely with your proposed architecture.

**Summary**: The paper proposes a dual-path attention mechanism integrated with UNet++ to improve the feature extraction and segmentation of thyroid nodules in ultrasound images. It aims to address the challenge of complex background and blurry boundaries common in medical imaging.

[**Link**](https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-024-01118-9): <https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-024-01118-9>

## DPAM-PSPNet

**Topic**: Thyroid Nodule Segmentation Using Dual-Path Attention Enhanced PSPNet

**Why it's a direct competitor?** Uses dual-path attention in a different architecture (PSPNet), serving as a comparison to attention-based U-Net models.

**Summary**: This model introduces dual-path attention into a PSPNet backbone for better global and local feature extraction. It achieves superior performance in segmenting thyroid nodules in noisy ultrasound images.

[**Link**](https://pubmed.ncbi.nlm.nih.gov/37439817/): <https://pubmed.ncbi.nlm.nih.gov/37439817/>

## TNPPD-Net

**Topic**: Precise Positioning of Ultrasound-Guided Fine-Needle Aspiration Biopsy of Thyroid Nodules

**Why it's a direct competitor?** It directly deals with biopsy alignment and segmentation of thyroid nodules, aligning with your research question of intent prediction.

**Summary**: TNPPD-Net is a novel framework that learns both segmentation and coordinate-based localization for biopsy guidance. It ensures precise alignment between predicted masks and actual biopsy puncture points.

[**Link**](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4763136): <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4763136>

## Swin-UNet

**Topic**: Swin Transformer Integrated UNet for Thyroid Nodule Segmentation

**Why it's a direct competitor?** Combines U-Net with Vision Transformer (Swin), offering a direct comparison for ViT-based approaches.

**Summary**: Swin-UNet fuses hierarchical Vision Transformers with U-Net’s decoder to capture global context and fine-grained details. It's designed to tackle edge preservation and spatial consistency in ultrasound segmentation.

[**Link**](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9751764/): <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9751764/>

## TRFE-Plus

**Topic**: Thyroid Region-Prior Feature Enhancement Network for Ultrasound Image Segmentation

**Why it's a direct competitor?** Introduces region-specific priors that improve segmentation accuracy, similar to attention enhancements in U-Net.

**Summary**: The TRFE-Plus network adds thyroid-specific spatial priors into the segmentation pipeline, helping the network focus on regions of interest. This reduces false positives in surrounding non-nodule areas.

[**Link**](https://doi.org/10.1016/j.compbiomed.2023.107442): <https://doi.org/10.1016/j.compbiomed.2023.107442>

## UNet + ResNet Encoder

**Topic**: Thyroid Nodule Segmentation Using UNet with ResNet Encoder

**Why it's a direct competitor?** A strong baseline U-Net variant using a deep encoder; critical for benchmarking U-Net performance.

**Summary**: This method integrates a ResNet encoder with a U-Net decoder to improve feature learning depth and contextual understanding. It achieves excellent accuracy across several thyroid segmentation metrics.

[**Link**](https://www.aimspress.com/article/doi/10.3934/mbe.2023202): <https://www.aimspress.com/article/doi/10.3934/mbe.2023202>

## Hybrid SegNet/U-Net with Despeckling

**Topic**: SegNet and UNet-Based Hybrid Framework with Despeckling for Ultrasound Segmentation

**Why it's a direct competitor?** Evaluates U-Net and SegNet architectures with pre-processing for improved ultrasound segmentation.

**Summary**: This hybrid model compares SegNet and U-Net after applying despeckling filters to enhance image clarity. It addresses speckle noise and blurry boundaries typical in ultrasound images.

[**Link**](https://link.springer.com/article/10.1007/s40477-023-00787-7): <https://link.springer.com/article/10.1007/s40477-023-00787-7>

## Multitask Detection + Segmentation

**Topic**: Multitask Learning for Detection and Segmentation of Thyroid Nodules

**Why it's a direct competitor?** Incorporates both segmentation and classification into a single pipeline, aligning with clinical interpretation.

**Summary**: A multitask network that jointly detects and segments thyroid nodules. It shares an encoder across both tasks to reduce redundancy and improve generalization.

[**Link**](https://www.medrxiv.org/content/10.1101/2023.06.15.23290671v1): <https://www.medrxiv.org/content/10.1101/2023.06.15.23290671v1>

## DSRU-Net

**Topic**: Deformable Pyramid Split-Attention Residual UNet (DSRU-Net) for Medical Imaging

**Why it's a direct competitor?** Enhances U-Net using deformable and pyramid attention, making it a top-performing U-Net-based method.

**Summary**: DSRU-Net adds deformable convolutions and split-attention modules into a residual U-Net. It enables multiscale learning and flexible kernel adaptation for better boundary recognition.

[**Link**](https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-023-00939-w): <https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-023-00939-w>

## P. Xu’s Improved U-Net

**Topic**: Improved UNet with BCE-Dice Loss and Convolutional Enhancements for Thyroid Nodule Segmentation

**Why it's a direct competitor?** Refines U-Net with better loss functions and convolution layers, improving edge precision.

**Summary**: This work modifies U-Net by optimizing its activation layers and introducing BCE-Dice loss. The improvements enhance boundary sharpness and segmentation reliability.

[**Link**](https://www.scipedia.com/public/Xu_2024a): <https://www.scipedia.com/public/Xu_2024a>

## Early Cascaded CNN (ICONIP 2018)

**Topic**: Cascaded CNN Framework for Early Thyroid Nodule Segmentation

**Why it's a direct competitor?** A foundational deep learning approach to thyroid segmentation before U-Net became standard.

**Summary**: Uses a two-stage cascaded CNN that first performs coarse segmentation and then refines the nodule boundaries. Serves as an early benchmark in this domain.

[**Link**](https://link.springer.com/chapter/10.1007/978-3-030-04221-9_28): <https://link.springer.com/chapter/10.1007/978-3-030-04221-9_28>

## CLIP-TNseg

**Topic**: CLIP-Based Vision Transformer for Weakly Supervised Thyroid Nodule Segmentation (CLIP-TNseg)

**Why it's a direct competitor?** Applies ViT + NLP (CLIP) for weakly supervised segmentation, relevant for clinical-aligned models.

**Summary**: CLIP-TNseg uses language-image embeddings from CLIP to guide segmentation from clinical text. The decoder combines this with spatial features to produce semantic segmentation masks.

[**Link**](https://arxiv.org/abs/2401.11892): <https://arxiv.org/abs/2401.11892>

## Asymmetric Pseudo-Label U-Net

**Topic**: Asymmetric Pseudo-Label Guided UNet for Coarse-to-Fine Segmentation

**Why it's a direct competitor?** Utilizes weak supervision and label refinement, relevant for settings with limited annotations.

**Summary**: Generates high-quality pseudo-labels from weak radiology reports and trains a U-Net using these iterative refinements. Ideal for improving performance without full supervision.

[**Link**](https://arxiv.org/abs/2404.06208): <https://arxiv.org/abs/2404.06208>

## MADLaP

**Topic**: Multi-modal Annotation via Deep Language and Pixel Model (MADLaP)

**Why it's a direct competitor?** Combines NLP and image segmentation to generate large-scale annotated thyroid datasets.

**Summary**: MADLaP integrates segmentation networks and NLP pipelines to auto-label nodules across datasets. It serves as a scalable pipeline for clinical dataset expansion.

[**Link**](https://arxiv.org/abs/2206.01667): <https://arxiv.org/abs/2206.01667>S

# 2. Vision Transformer Usage in Competitors:

Only two of the 14 competitor papers incorporate Vision Transformer (ViT) architectures:

1. Swin-UNet  
2. CLIP-TNseg

## Swin-UNet

This model integrates the Swin Transformer as an encoder into a U-Net architecture. It benefits from hierarchical self-attention to model long-range dependencies while maintaining spatial resolution through a U-Net-style decoder. The method is effective for preserving edge information and spatial context.

## CLIP-TNseg

CLIP-TNseg uses Vision Transformer (ViT) embeddings from the CLIP model to guide weakly supervised segmentation of thyroid nodules. These embeddings are aligned with clinical text, allowing the model to segment semantic regions even without pixel-perfect annotations.

# 3. Methodologies of All Competitor Papers:

• **DPAM-UNet++:** Enhances the U-Net++ backbone with a dual-path attention mechanism to better extract both low-level and high-level features, boosting accuracy in nodule segmentation.

• **DPAM-PSPNet:** Integrates dual-path attention into PSPNet’s encoder to improve contextual understanding and spatial localization of thyroid nodules.

• **TNPPD-Net:** Uses a U-Net-like architecture with an added coordinate prediction head to precisely align predicted nodules with biopsy puncture intent.

• **Swin-UNet:** Leverages Swin Transformer as encoder and a U-Net-style decoder to model long-range dependencies and preserve spatial hierarchies in ultrasound segmentation.

• **TRFE-Plus:** Applies region-prior attention features to guide the segmentation network’s focus toward the thyroid region, enhancing boundary localization.

• **UNet + ResNet Encoder:** Incorporates a pre-trained ResNet encoder in a U-Net to capture more abstract feature representations from thyroid ultrasound images.

• **Hybrid SegNet/U-Net with Despeckling:** Combines despeckling preprocessing with a hybrid U-Net and SegNet model to improve image clarity and segmentation accuracy.

• **Multitask Detection + Segmentation:** Uses a shared encoder for detection and segmentation tasks with separate decoders to simultaneously identify and delineate thyroid nodules.

• **DSRU-Net:** Incorporates deformable convolutions and pyramid split-attention within a residual U-Net framework to enhance flexibility and multiscale feature extraction.

• **P. Xu’s Improved U-Net:** Improves standard U-Net with additional convolution layers, activation tweaks, and BCE-Dice loss for sharper nodule edge detection.

• **Early Cascaded CNN (ICONIP 2018):** A cascaded CNN pipeline progressively refines segmentation results through multiple connected stages for better final boundary definition.

• **CLIP-TNseg:** Fuses Vision Transformer (CLIP) text-image embeddings with CNN decoders to perform weakly supervised segmentation based on report-aligned cues.

• **Asymmetric Pseudo-Label U-Net:** Guides U-Net segmentation using refined pseudo-labels generated from coarse annotations, improving accuracy on weakly labeled data.

• **MADLaP**: Integrates deep learning segmentation models with NLP-based annotation extraction to auto-label thyroid nodules using multi-modal data.

# 4. Results from All Competitor Papers

• **DPAM-UNet++:** Dice: 0.83, IoU: 0.745, AUC: 0.921

• **DPAM-PSPNet**: Dice: 0.9213, mIoU: 0.8675

• **TNPPD-Net:** Dice: ~0.92, IoU: ~0.86, Accuracy: 99.3%

• **Swin-UNet:** Dice: ~0.78, IoU: ~0.71

• **TRFE-Plus:** Dice: ~0.84, IoU: ~0.78

• **UNet + ResNet Encoder:** Dice:0.842, IoU: 0.755, Accuracy: 97.2%

• **Hybrid SegNet/U-Net with Despeckling**: Dice: ~0.88, IoU: ~0.87

• **Multitask Detection + Segmentation:** Dice: ~0.85, IoU: ~0.82

• **DSRU-Net:** Dice: 0.925, IoU: 0.858

• **P. Xu’s Improved U-Net:** Dice: 0.9062, Accuracy: 96.8%

• **Early Cascaded CNN (ICONIP 2018):** Dice: ~0.81, IoU: ~0.75

• **CLIP-TNseg:** Dice**:** 0.89, mIoU: 0.86

• **Asymmetric Pseudo-Label U-Net:** Dice: ~0.84, IoU: ~0.78

• **MADLaP:** Dice: ~0.86, IoU: ~0.81

# 5. Comparison: Vision Transformer vs U-Net Models

ViT-based models like Swin-UNet and CLIP-TNseg offer significant advantages in modeling long-range dependencies and integrating contextual or semantic features. This makes them ideal for edge preservation and weakly supervised scenarios. However, U-Net and its variants remain dominant in clinical deployments due to their interpretability, training efficiency, and compatibility with smaller datasets.

**Summary of performance:**

• **Swin-UNet:** Dice ~0.78 | IoU ~0.71

• **CLIP-TNseg**: Dice ~0.89 | IoU ~0.86

• **DPAM-UNet++:** Dice ~0.83 | IoU ~0.745

• **UNet + ResNet:** Dice ~0.842 | IoU ~0.755

• **TNPPD-Net:** Dice ~0.92 | IoU ~0.86 (also targets biopsy alignment)