**UNet with Fractal and Fractional Attention Mechanisms for Accurate Localization and Segmentation of Lung Nodules**

**ALL Tables**

**Table 1:** Summary of the number of CT scans and patients in the LIDC-IDRI datasets used in this study.

|  |  |
| --- | --- |
| **Description** | **LIDC-IDRI** |
| Total number of patients | 1010 |
| Total number of CT scans | 1018 |
| Final Nodules | 848 (442 benign, 406 malignant) |
| Origin | Seven academic instutions. |
| CT Slice width | 0.45 to 5.00 mm |
| Annotation | Radiologists Four experienced thoracic radiologists. |
| Barred Nodules | Rating median of 3 |
| Experts annotation | Median of radiologists' ratings (1-5 scale) |
| Url of LIDC-IDRI | [**https://www.cancerimagingarchive.net/collection/lidc-idri/**](https://www.cancerimagingarchive.net/collection/lidc-idri/) |
| Url of LUNA 16 | [**https://luna16.grand-challenge.org/Data/**](https://luna16.grand-challenge.org/Data/) |

**Table 2**. 3X3 Fractal-fractional integral mask

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

**Table 3.** UNet-FFA model parameters

|  |  |
| --- | --- |
| **Parameter** | **Configuration** |
| Learning rate | 0.001 |
| Epochs | 100 |
| Optimizer | Adam |
| Activation function | ReLU |
| Batch size | 32 |

**Table 4.** Datasets applied to various methods.

|  |  |  |
| --- | --- | --- |
| **Methods** | **Year** | **Datasets** |
| IOMT assisted modified UNet [27] | 2021 | LIDC-IDRI |
| DMC U-Net [28] | 2023 | LIDC-IDRI |
| SMR Unet | 2023 | LIDC-IDRI |
| UNet-based neural network [29] | 2023 | CT slices (Private) |
| ACX U-Net [30] | 2023 | LIDC-IDRI,LUNA 16 |
| 3D Unet [31] | 2024 | LUNA 16 |
| HMSAM-Unet | 2024 | LUNA 16 |
| MAST Unet [32] | 2024 | LIDC-IDRI |

**Table 5**. Comparison of segmentation performance for different UNet and UNet models with attention mechanism models for the LUNA 16 dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Methods** | **Year** | **Dice-coefficient** | **IoU** | **Precision** | **Recall** |
| ACX U-Net | 2023 | 0.98 | 0.97 | 0.98 | 0.98 |
| 3D UNet | 2024 | 0.84 | 0.74 | - | - |
| HMSAM UNet muti-scale attention model | 2024 | 0.98 | 0.97 | 0.99 | 0.99 |
| MUS-Net | 2025 | 0.87 | 0.78 | - | - |
| 3D Residual network ResNet 50 | 2025 | 0.96 | 0.90 | 0.95 | 0.97 |
| **UNet-FFA** | **2025** | **0.98** | **0.97** | **0.97** | **0.96** |

**Table 6**. Comparison of segmentation performance for different UNet and UNet models with attention mechanism models for the LIDC-IDRI dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Methods** | **Year** | **Dice-coefficient** | **IoU** | **Precision** | **Recall** |
| IOMT-assisted modified UNet [27] | 2021 | 0.85 | 0.74 | - | 0.86 |
| DS-Trans UNet [33] | 2022 | 0.92 | 0.86 | 0.90 | 0.94 |
| DMC U-Net [28] | 2023 | - | 0.83 | 0.91 | 0.91 |
| UNet-based neural network [29] | 2023 | 0.93 | 0.90 | 0.93 | 0.95 |
| SMR Unet with self attention | 2023 | 0.91 | 0.86 | 0.93 | 0.92 |
| RAD Unet with channel attention mechanism | 2023 | - | 0.87 | 0.94 | 0.92 |
| TPFR-net U Shaped dual attention | 2023 | 0.91 | - | 0.92 | - |
| ResDSda UNet with self attention | 2023 | 0.86 | 0.76 | 0.87 | 0.86 |
| MAST UNet | 2024 | 0.86 | 0.86 | - | - |
| ParaU-Net [38] | 2024 | 92.16 | 87.15 | - | - |
| CAFU-Net | 2025 | 0.91 | 0.86 | - | - |
| **UNet-FFA** | **2025** | **0.98** | **0.97** | **0.96** | **0.95** |

**Table 7 Comparison of UNetFFA with Transformer models**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Methods** | **Year** | **Dice-coefficient** | **Strengths** | **Limitations** | **Why UNet-FFA is better** |
| DSNet [46] | 2024 | 0.93 | Low latency, fewer parameters | Poor contextual reasoning | Fractional attention captures long-term memory; fractal attention manages scale invariance without excessive architecture. |
| Multi-scale transformer [47] | 2025 | 0.94 | Intense worldwide focus, effective feature representation. | Significant expense in computation and memory; necessitates extensive datasets. | UNet-FFA offers both global and local attention while requiring considerably fewer resources. |
| Hybrid UNet with Visual Transformer [48] | 2025 | 0.81 | A strong global emphasis and successful feature representation. | High costs in processing and memory are required; this demands large datasets. | UNet-FFA utilizes significantly fewer resources while delivering both local and global attention. |
| Two-stage UNet [49] | 2025 | 0.89 | Low latency, fewer parameters | Poor contextual reasoning | Fractional attention captures long-term memory; fractal attention manages scale invariance without excessive architecture. |
| Multi-scale feature fusion transformer UNet [50] | 2025 | 0.80 | There is a strong global focus on effective feature representation. | Significant expenses for processing and memory are necessary; this requires substantial datasets. | UNet-FFA requires significantly fewer resources while delivering attention on both a local and global scale. |

**Table 8**. *P*-values from a paired *t*-test between our models and other methods

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Metrics** | **3D UNet** | **IOMT** | **DS-Trans UNet** | **UNet** | **SMR UNet** | **TPR-Net** | **MAST UNet** | **ParaUNet** | **UNet-FFA** |
| ***P*-value** | 0.011 | 0.177 | 0.007 | 0.032 | 0.004 | 0.004 | 0.016 | 0.004 | **0.002** |

**Table 9** Ablation study results for LIDC-IDRI dataset with different connections

|  |  |  |  |
| --- | --- | --- | --- |
| **Models** | **Dice coefficient** | **IoU** | **Comments** |
| UNet | 0.84 | 0.75 | UNet is designed to capture and represent local context within data, only effectively |
| UNet with Attention Mechanism | 0.85 | 0.76 | Focus on local attention for feature channels or pixels only. But UNet-FFA focuses on global and local attention for pixels. |
| UNet-FFA on encoder side only | 0.89 | 0.83 | Using attention in the decoder may result in poorer segmentation of fine details, such as sharp edges and small blobs. |
| UNet-FFA on the Decoder side only | 0.90 | 0.82 | May perform less effectively on very intricate textures in comparison to complete attention integration. |
| UNet-FFA without skip connections | 0.87 | 0.81 | The local spatial detail may be lost, resulting in the potential loss of small and fine structures. |
| **UNet-FFA (LIDC-IDRI Dataset)** | **0.98** | **0.97** | Compared to standard attention, our proposed model performs tasks on irregular boundaries, low contrast nodules, and multi-scale targets. |

**Table 10** Ablation study results for the LUNA16 dataset with different connections

|  |  |  |  |
| --- | --- | --- | --- |
| **Models** | **Dice coefficient** | **IoU** | **Comments** |
| UNet | 0.82 | 0.78 | UNet is structured to capture and portray local context found within data exclusively efficiently. |
| UNet with Attention Mechanism | 0.81 | 0.77 | UNet with attention may miss or blur, but FFA preserves well. |
| UNet-FFA on the encoder side only | 0.91 | 0.83 | The decoder acquires enhanced features; however, it does not further refine them. |
| UNet-FFA on the Decoder side only | 0.92 | 0.85 | There is a lack of enhancement in the encoder features, as the initial feature extraction process may overlook subtle patterns. |
| UNet-FFA without skip connections | 0.89 | 0.85 | The lack of skip connections in a neural network architecture can harm the accuracy of edge detection. |
| **UNet-FFA (LUNA 16 Dataset)** | **0.98** | **0.97** | UNet-FFA excels in edge preservation, reducing false positives and capturing complex shapes. |