



Seminar - Fall 2023

Medical Image Segmentation for Realizing Human Digital Twin

Implementation and Results

SYED HASNAIN RAZA SHAH

Outline

1. Recap

2. Preliminary Experiments

3. Results

Outline

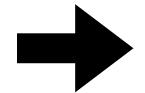
1. Recap

2. Preliminary Experiments

3. Results

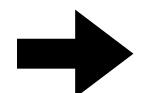
Introduction

Problem



- To identify precise structure of bones and muscles from CT/MRI images in order to build human digital twin

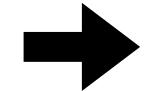
Motivation



- Lack of research focusing on Segmentation of Bones and Muscles
- Advancing Healthcare
- Personalized treatment strategies, optimizing outcomes and reducing risks.
- Assisment in surgical planning, leading to safer and more efficient surgical procedures.
- Segmentation-driven digital twins are at the forefront of biomedical research and technology.

Recap

**Research
Goals**



- Understanding State-of-the-art techniques for medical image segmentation
- Acquiring the latest Vision Transformer-based expertise
- Development of segmentation model for bones and muscles on Visible Korean

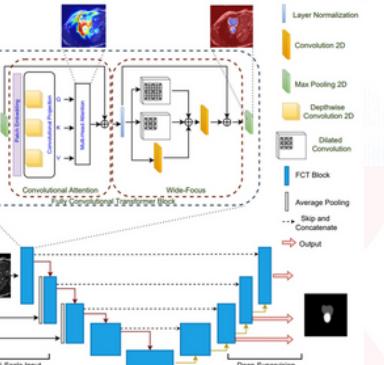
Recap

Fully Convolutional Transformer

Neither Transformer-CNN hybrid nor Transformer-UNet

Uses *FCT layer* as building block

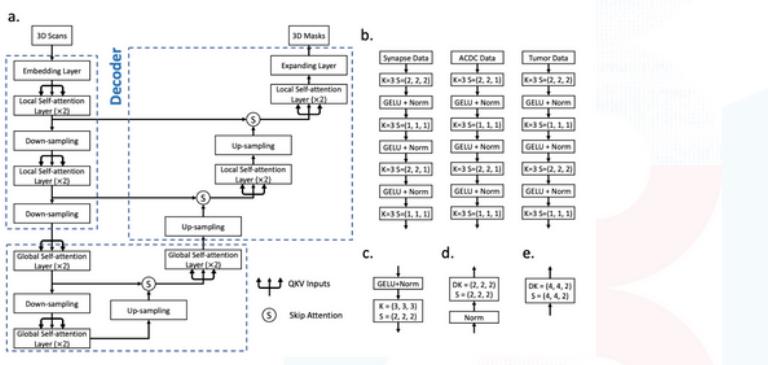
- consists of convolutional layers followed by Gelu activation function
- convolutional attention module replacing linear projection with Depthwise-Convolutions, removes positional encoding
- wide focus module contains dilated convolutions and convolutional layer for feature aggregation



"The fully convolutional transformer for medical image segmentation," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023.

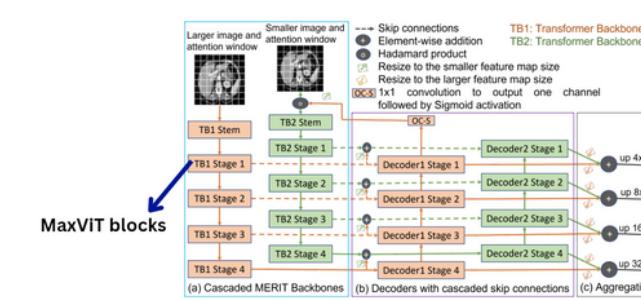
nnFormer

nnFormer



"nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer," in IEEE Transactions on Image Processing, vol. 32, pp. 4036-4045, 2023.

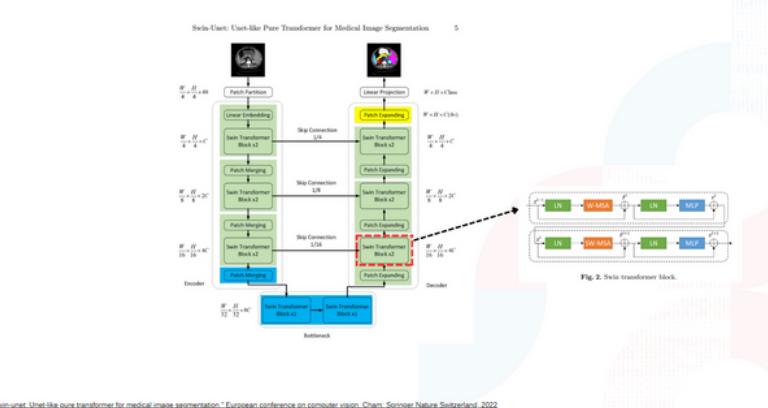
Multi-scale Hierarchical Vision Transformer (MERIT)



- combines MaxViT and Cascaded Decoders
- captures both multi-scale and multi-resolution features

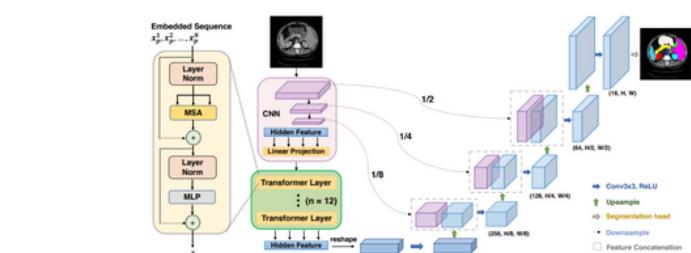
"Multi-scale Hierarchical Vision Transformer with Cascaded Attention Decoders for Medical Image Segmentation," MDL, 2023

SwinUnet



"SwinUnet: Unit-like Pure Transformer for Medical Image Segmentation," European conference on computer vision, Cham, Springer Nature Switzerland, 2022.

TransUnet



"TransUnet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.

Visible Korean

- Full-body scans of a Korean man and a woman, and the corresponding CT, MRI, and segmentation masks
- The segmentation mask consists of a total of 13 major categories
 - MRI: 4mm (256x256, 12bit)
 - CT : 1mm (512,512, 12bit, 1702 in total)
 - BodySlice: 1mm (2048x1216, 24bit, 1702 in total)



Our Initial Approach

Medical Image Segmentation on Automatic Cardiac Diagnosis Challenge (ACDC)

Rank	Model	Avg DSC ↑ Paper	Code	Result	Year	Tags
1	FCT	93.02	The Fully Convolutional Transformer for Medical Image Segmentation			2022
2	MERIT	92.32	Multi-scale Hierarchical Vision Transformer with Cascaded Attention Decoding for Medical Image Segmentation			2023
3	nnFormer	92.06	nnFormer: Interleaved Transformer for Volumetric Segmentation			2022
4	TransCASCADE	91.63	Medical Image Segmentation via Cascaded Attention Decoding			2022
5	PVT-CASCADE	91.46	Medical Image Segmentation via Cascaded Attention Decoding			2022
6	SwinUnet	90.00	Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation			2021
7	TransUNet	89.71	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021
8	MISSFormer	87.9	MISSFormer: An Effective Medical Image Segmentation Transformer			2021
9	R50-ViT-CUP	87.57	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021
10	R50-AttnUNet	86.75	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021

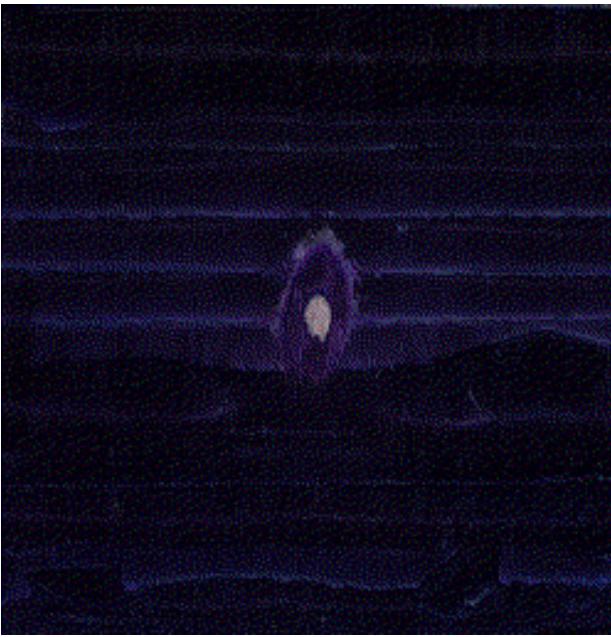
Outline

1. Recap

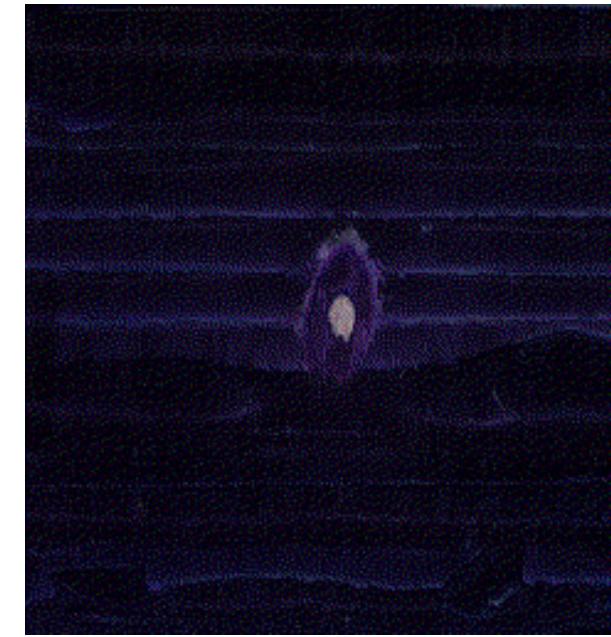
2. Preliminary Experiments

3. Results

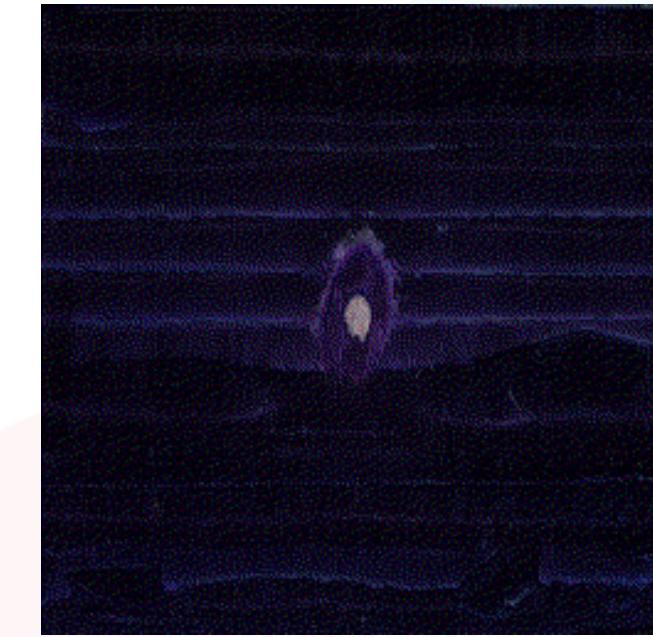
Data Preprocessing



VK-Male



VK-Male with Muscular Mask



VK-Male with Skeletal Mask

- Using body slice images of a full-body male, male head, female pelvis and bone/muscle segmentation masks.
- Bone and muscle segmentation masks consist of 191 and 265 subcategories, respectively.
- For each consecutive slice, a musculoskeletal segmentation mask was created by integrating bone subcategories and muscle subcategories respectively.

Experimental Setting

► Dataset

- **Visible Korean [Body Slice + Head + Female Pelvis] (2186 in total)**
- **Preprocessed GT Masks**
- **8:2 (train:validation) and 14 Images for quality check**

► Performance Measure

- **Dice coefficient**
- **HD95**

$$DSC = \frac{2|GT \cap PT|}{|GT| + |PT|} \quad (1)$$

$$d_H(X, Y) = \max\{d_{XY}, d_{YX}\} = \max \left\{ \max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y) \right\} \quad (2)$$

► GPU


NVIDIA x 2
A100
80 GB

Preliminary Experiments

Model	Input	LR	LR Reduce	batch size	optimizer	loss	epoch
FCT	256	1e-3	T	2	Adam	dice, cross entropy	50 (warmup) + 250 (training)
MERIT	256	1e-4	F	75	AdamW	dice, cross entropy	400
SwinUnet	256	1e-2	T	150	SGD	dice, cross entropy	300
TransUnet	256	1e-2	T	150	SGD	dice, cross entropy	300
PVT-Cascade	256	1e-4	F	150	AdamW	dice, cross entropy	300
Trans-Cascade	256	1e-4	F	150	AdamW	dice, cross entropy	300

Preliminary Experiments

Hyperparameter Tuning

Model	Input	LR	LR Reduce	batch size	optimizer	loss	epoch
FCT	256	5e-4	T	2	Adam	dice, cross entropy	70 (warmup) + 500 (training)
MERIT	256	3e-4	F	75	AdamW	dice, cross entropy	390
SwinUnet	256	5e-2	F	150	SGD	dice, cross entropy	1540
TransUnet	256	7e-2	F	150	SGD	dice, cross entropy	2320
PVT-Cascade	256	9e-4	F	150	AdamW	dice, cross entropy	300
Trans-Cascade	256	4e-4	F	150	AdamW	dice, cross entropy	300

Outline

1. Recap

2. Preliminary Experiments

3. Results

Results

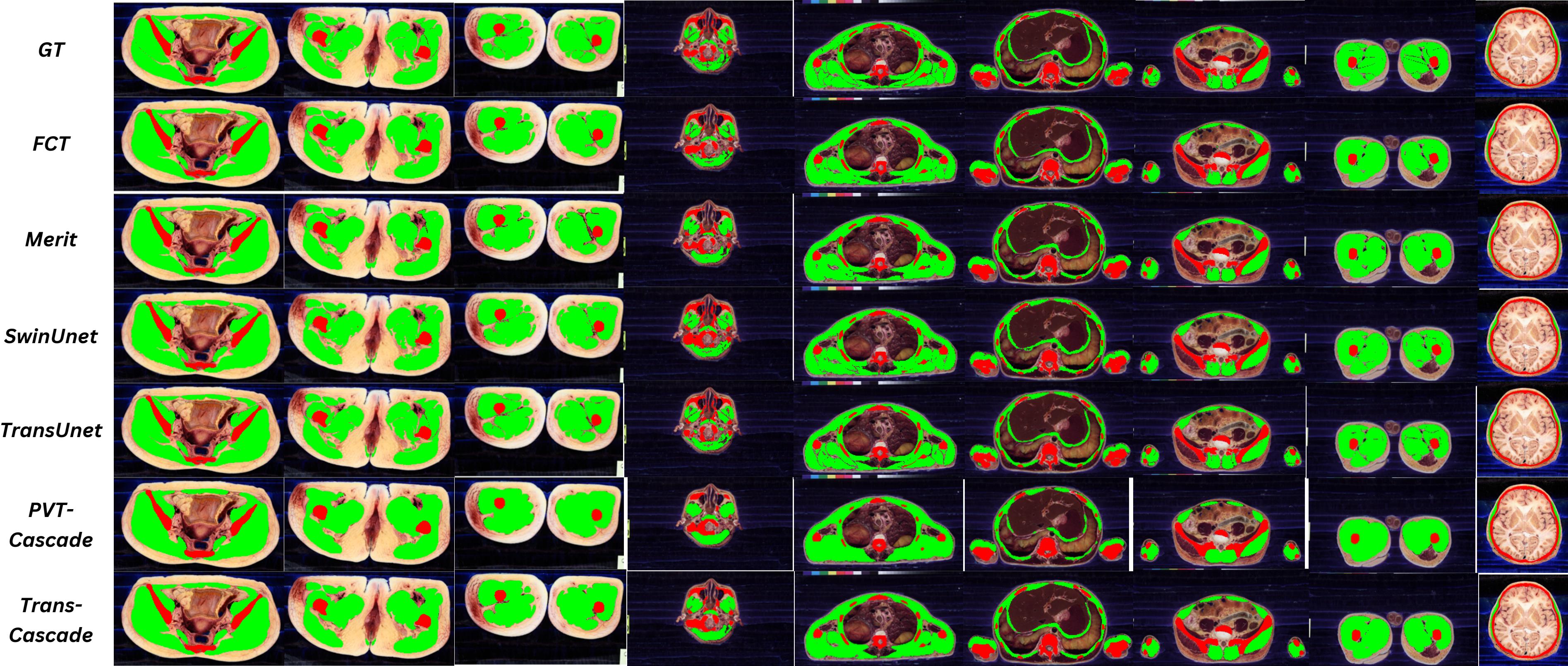
Quantitative

Model	Skeletal Dice	Muscle Dice	Skeletal HD95	Muscle HD95	Mean Dice	Mean HD95
FCT	0.9379	0.932	1	1	0.93495	1
MERIT	0.9146	0.8451	1.9094	5.6755	0.87985	3.79245
SwinUnet	0.9157	0.853	1.5398	4.3377	0.88435	2.93875
TransUnet	0.946	0.8903	1.3244	3.0362	<u>0.91815</u>	<u>2.1803</u>
PVT-Cascade	0.8874	0.8117	4.4079	11.774	0.84955	8.09095
Trans-Cascade	0.9149	0.8496	2.5578	8.7865	0.88225	5.67215



Results

Qualitative



Research Outcome

휴먼 디지털 트윈 개발을 위한 Visible Korean 이미지 자동 세그멘테이션 모델*

Syed Hasnain Raza Shah^{1,2}, 양동현¹, 이경하^{1,2,*}

¹과학기술연합대학원대학교 (UST), ²한국과학기술정보연구원 (KISTI)

hasnain@kisti.re.kr, yangdonghun3@kisti.re.kr, kyongha@kisti.re.kr

Automated Segmentation Model of Visible Korean Images for Human Digital Twin

Syed Hasnain Raza Shah^{1,2}, Donghun Yang¹, Kyong-Ha Lee^{1,2}

¹University of Science and Technology (UST),

²Korea Institute of Science and Technology Information (KISTI)

요약

인간의 해부학적 특성과 생리학적 상태를 디지털 공간에 재현하는 휴먼 디지털 트윈은 최근 의료 분야에서 많은 관심을 받고 있다. 정밀한 휴먼 디지털 트윈 개발을 위해서는 인간의 신체를 구성하는 다양한 요소들을 정밀하게 모델링해야 한다. 특히 인체의 구조를 결정하는 뼈와 근육 구조를 정밀하게 모델링 하는 것이 중요하다. 이를 위해 고품질의 다양한 해부학적 이미지가 필요하며, 해당 이미지에서 인체의 구조적인 정보를 효과적으로 추출할 수 있는 자동 세그멘테이션 기술의 개발이 필수불가결하다. 이에 본 연구는 한국과학기술정보연구원 (KISTI) 이 보유한 대규모 고품질 해부학 이미지 데이터셋인 Visible Korean 과 최근 의료 이미지 세그멘테이션 분야에서 가장 우수한 성능을 보이는 Fully Convolutional Transformer (FCT) 모델을 기반으로 근골격 구조 추출을 위한 자동 세그멘테이션 모델에 관한 연구를 수행하였다. 실험 결과 FCT 모델은 Visible Korean 이미지 데이터셋에서 근육 및 뼈 구조를 각각 Dice Similarity Coefficient (DSC) 0.9890 및 0.9973의 높은 정확도로 추출하였다. 본 연구 결과는 휴먼 디지털 트윈 기술 발전에 크게 이바지할 것으로 기대된다.

1. 서론

인간의 해부학적 특성과 생리학적 상태를 디지털 공간에 재현하는 휴먼 디지털 트윈은 최근 의료 분야에서 많은 관심을 받고 있다. 이러한 휴먼 디지털 트윈은 실제 환자의 생체 정보를 정밀하게 분석하여 진단, 치료, 수술 등의 의료행위를 시뮬레이션하는 데 활용될 수 있으며, 이를 통해 의료진은 환자 개개인의 생체 정보를 더욱 정확하게 이해하여 환자별 맞춤형 치료 및 수술 계획을 수립할 수 있다. 따라서 정밀하고 신뢰성 높은 휴먼 디지털 트윈의 개발은 의료분야에서 중요한 연구 주제이다 [1].

정밀한 휴먼 디지털 트윈 개발을 위해서는 인간의 신체를 구성하는 다양한 요소들을 정밀하게 모델링해야 한다. 신체의 다양한 요소를 중 뼈와 근육으로 구성되는 근골격 구조는 인체의 구조를 결정하는 핵심 구성요소이다. 뼈는 인체를 지탱하고 보호하는 역할을 하며, 근육은 뼈와 연결되어 수축과 이완에 따라 인체의 움직임을 제어한다. 휴먼 디지털 트윈이 실제 인체와 유사한 움직임을 정밀하게 재현하기 위해서는 이러한 근골격 구조를 정밀하게 모델링하는 것이 중요하다 [2].

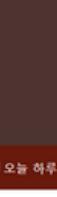
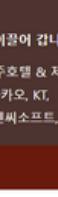
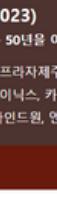
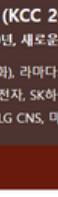
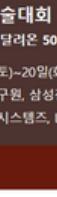
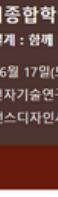
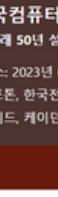
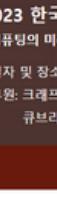
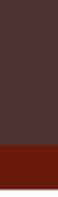
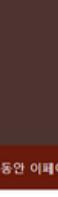
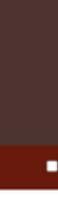
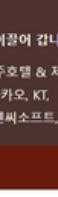
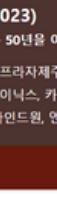
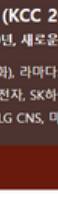
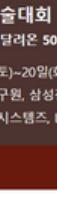
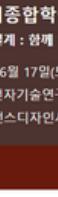
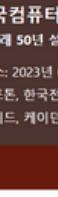
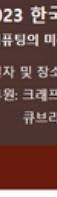
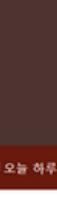
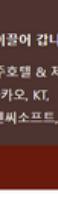
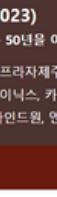
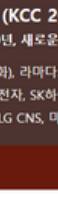
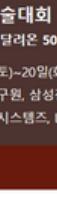
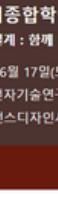
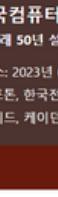
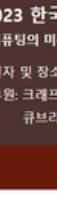
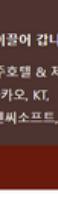
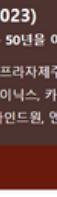
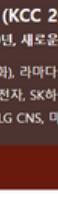
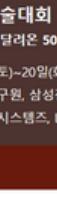
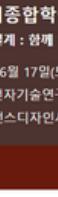
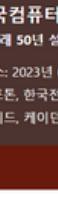
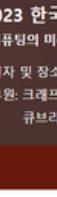
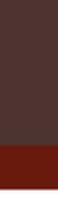
근골격 구조를 포함하는 신체의 다양한 구성요소들을 정밀하게 모델링하기 위해서는 CT, MRI, X-ray 등 고품

질의 다양한 해부학적 이미지를 필요하며, 해당 이미지를 효율적으로 구조 정보를 추출할 수 있는 자동 세그멘테이션 기술이 필수적이다. 자동 세그멘테이션 기술은 이미지에서 뼈, 근육 등의 다양한 구조들을 자동으로 분리하고 추출할 수 있는 기술이며, 딥러닝과 같은 인공지능 기술의 발전으로 최근 더욱 정확한 구조정보 추출 성능을 보인다 [3].

따라서 본 연구는 한국과학기술정보연구원 (KISTI) 이 보유하고 있는 대규모 고품질 해부학 영상이미지 데이터셋인 Visible Korean 을 기반으로 근골격 구조 추출을 위한 세그멘테이션 모델에 관한 연구를 수행하였다 [4]. 먼저 세그멘테이션 모델 학습을 위해 Visible Korean 데이터셋에 대한 전처리를 수행하였다. 전처리된 데이터를 기반으로 최근 의료 이미지 세그멘테이션 분야에서 가장 우수한 성능을 보이는 Fully Convolutional Transformer (FCT) 모델을 학습하여 근골격 구조 자동 세그멘테이션 모델을 구현하였으며, 그 성능을 검증하였다 [5]. 본 연구의 결과는 의료분야에서 근골격 구조 분석 및 진단에 다양하게 활용 될 수 있으며, 정밀한 휴먼 디지털 트윈 개발에 기여할 수 있을 것으로 기대된다.

2. Visible Korean 데이터셋

본 연구는 Visible Korean 데이터셋을 활용하여 근골격 구조 자동 세그멘테이션 모델을 개발하는 것을 목표로



Future Works

- Design our own architecture and compare the performance with existing methods
- Design Unet based on latest versions of vision transformer backbones
- Multi parallel encoders (e.g Swin, ViT, FCT)
- More CT/MRI data is expected

Thank You!

