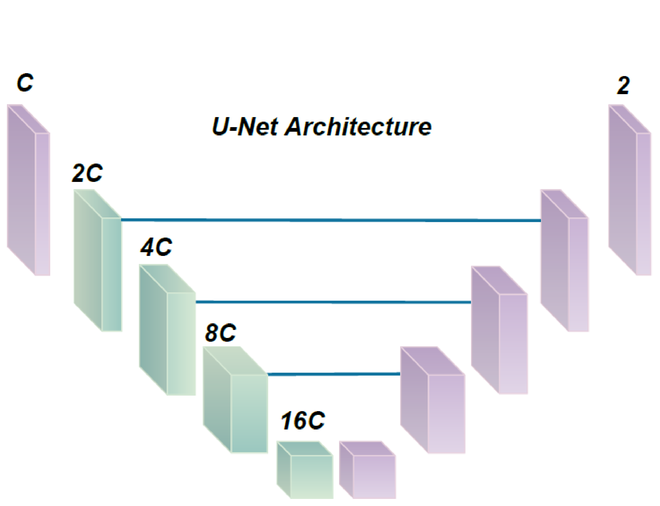
U-Net vs Transformer: Is U-Net outdated in Medical Image registration?

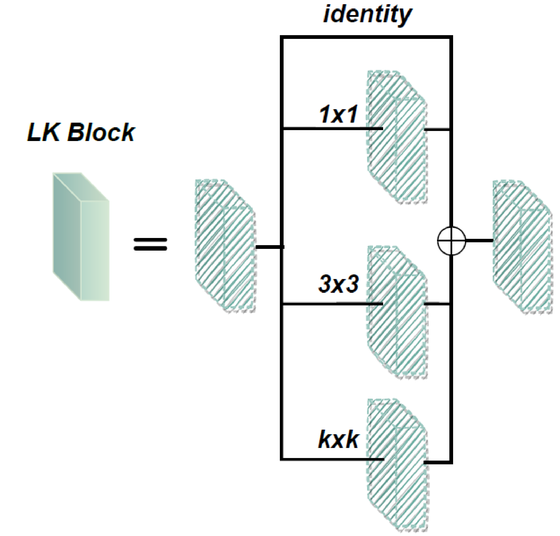
Xi Jia, Joseph Bartlett, Tianyang Zhang, Wenqi Lu, Zhaowen Qiu, Jinming Duan

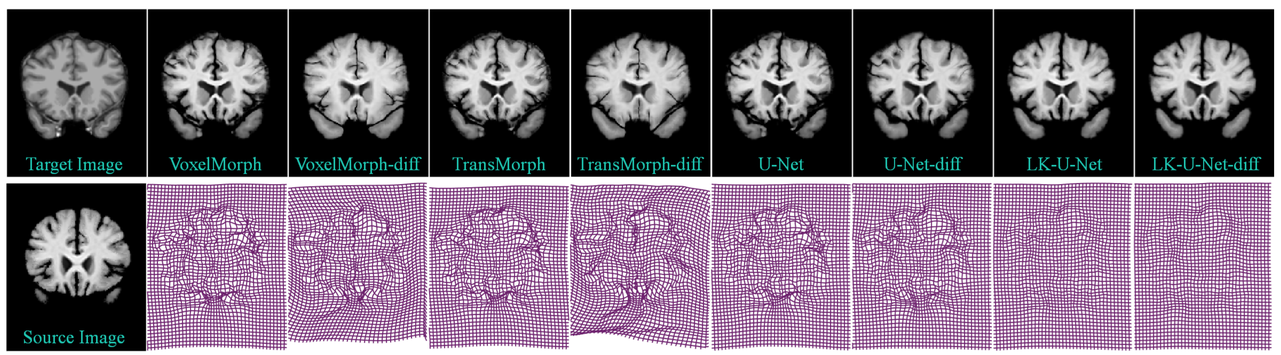
UNIVERSITY OF BIRMINGHAM

#1 Introduction

- **Optimization based Methods**
  iterative pair-wise optimization; slow but accurate.
- **Convolutional U-Net based Models (VoxelMorph)**
  comparable with **optimization methods** but much faster.
- **Transformer (hybrid) based Models (TransMorph)**
  long-range dependencies; more accurate than **VoxelMorph**.



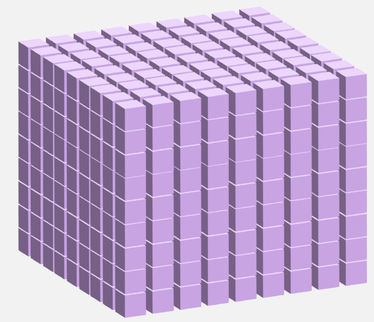




#2 Method

Is transformer all we need?

- We find the voxel displacements (sphere) are much smaller than the volume size (cube).
- We doubt the necessity of adopting a transformer to model long-range dependencies.



Simple Technique: Large Kernel U-Net

- We use a 5-layer backbone U-Net with 3×3 kernels.
- Following the *RepVGG* and *RepLK-Net*, we replace the 3x3 kernel with four parallel kernels, which are identity, 1×1, 3×3, and $k \times k$ large kernel, respectively.

#3 Results

Evaluated on the IXI and OASIS 3D brain datasets.

- LKU-Net outperforms optimization based methods.
- LKU-Net outperforms VoxelMorph.
- LKU-Net outperforms TransMorph.
  - 1.12% of the parameters and 10.8% of the mult-adds.

Methods	Dice	%|J|<0	Parameters	Mult-Adds(G)
deedsBCV	0.733	0.147±0.050	-	-
VoxelMorph-2	0.726	1.522±0.336	301,411	398.81
Backbone U-Net	0.727	1.524±0.353	279,086	58.73
TransMorph	0.746	1.579±0.328	46,771,251	657.64
LKU-Net	0.752	0.023±0.018	522,302	71.00