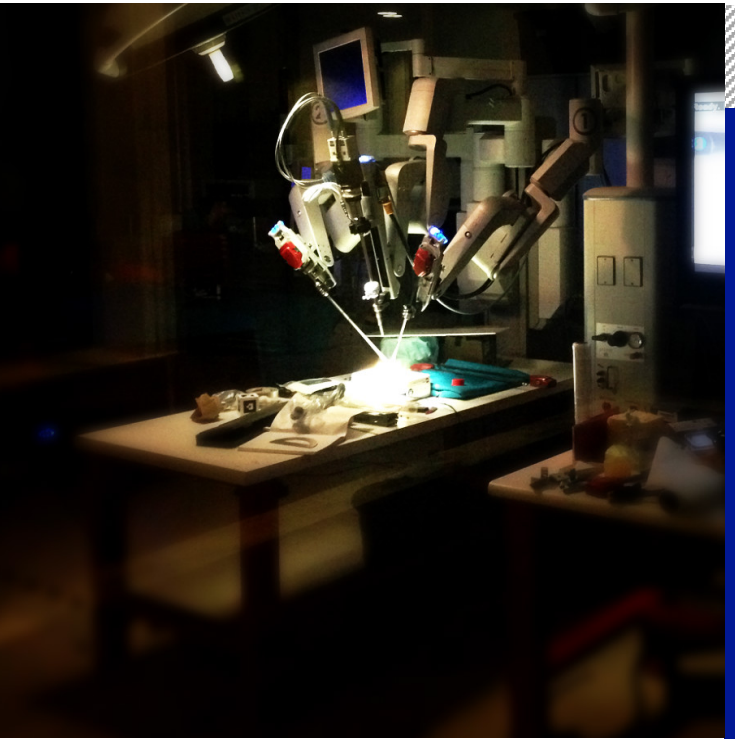


Learning to See Forces:

Surgical Force Prediction with RGB-Point Cloud Temporal Convolutional Networks

Cong Gao, Xingtong Liu, Mathias Unberath, Michael Peven, Austin Reiter

The Johns Hopkins University

JOHNS HOPKINS
MALONE CENTER *for*
ENGINEERING *in* HEALTHCARE

LABORATORY FOR
Computational
Sensing + Robotics

Background

- Haptic feedback is one major limitation of current Robotic Surgical Systems[1]
- Obtaining forces during surgery is difficult
- Long-studied problem (most are hardware-based solutions)
  - Hardware is: expensive, difficult to sterilize, etc.

[1] Lee, Chiwon, etal. "A grip force model for the da Vinci end-effector to predict a compensation force." Medical & biological engineering & computing 53.3 (2015): 253-261.

How About Visual Cues?

- Surgeons rely on visual cues to infer forces from tissue deformation in the absence of physical haptic feedback.[2]
  - Eg. Stereo Camera in the current da Vinci Surgical System.
- Computer vision has been used to measure the deformed object and recover the applied force from linear elasticity equations.
- We propose to mimic this capability using deep learning to infer forces from video.

[2] DiMaio, Simon, Mike Hanuschik, and Usha Kreaden. "The da Vinci surgical system." Surgical Robotics. Springer, Boston, MA, 2011. 199-217.

Dataset Collection

Video Data Collection:

- Kinect2 RGB+D Camera

RGB Image

Depth Image

crop

Region of Interest

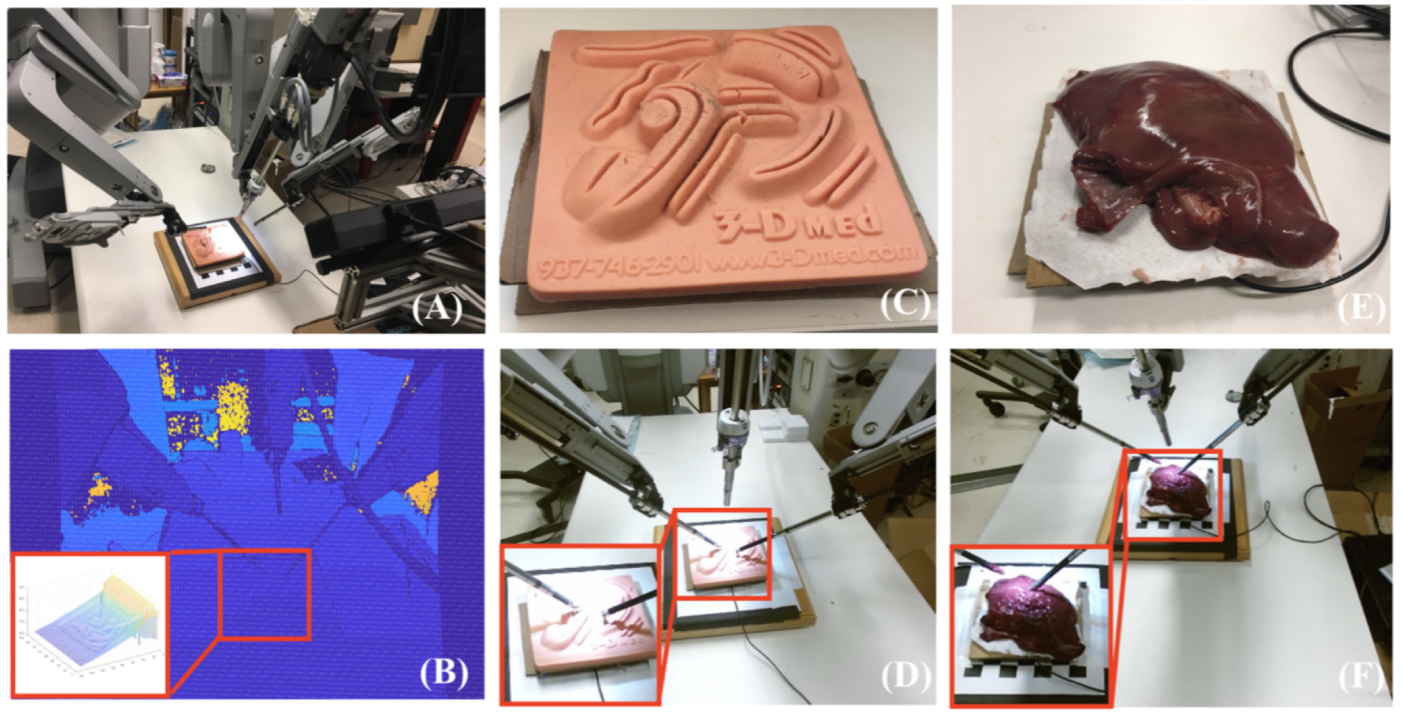


Fig. 2. (A) The experimental environment. (B) The depth image from the Kinect2 camera. (C) The phantom used in the experiment. The force sensor is placed under the phantom. (D) RGB image of the phantom. (E) A fresh piece of pig liver. (F) RGB image of the liver.

- Video Stream flows at 30 fps
- Synchronized to be within 10 ms

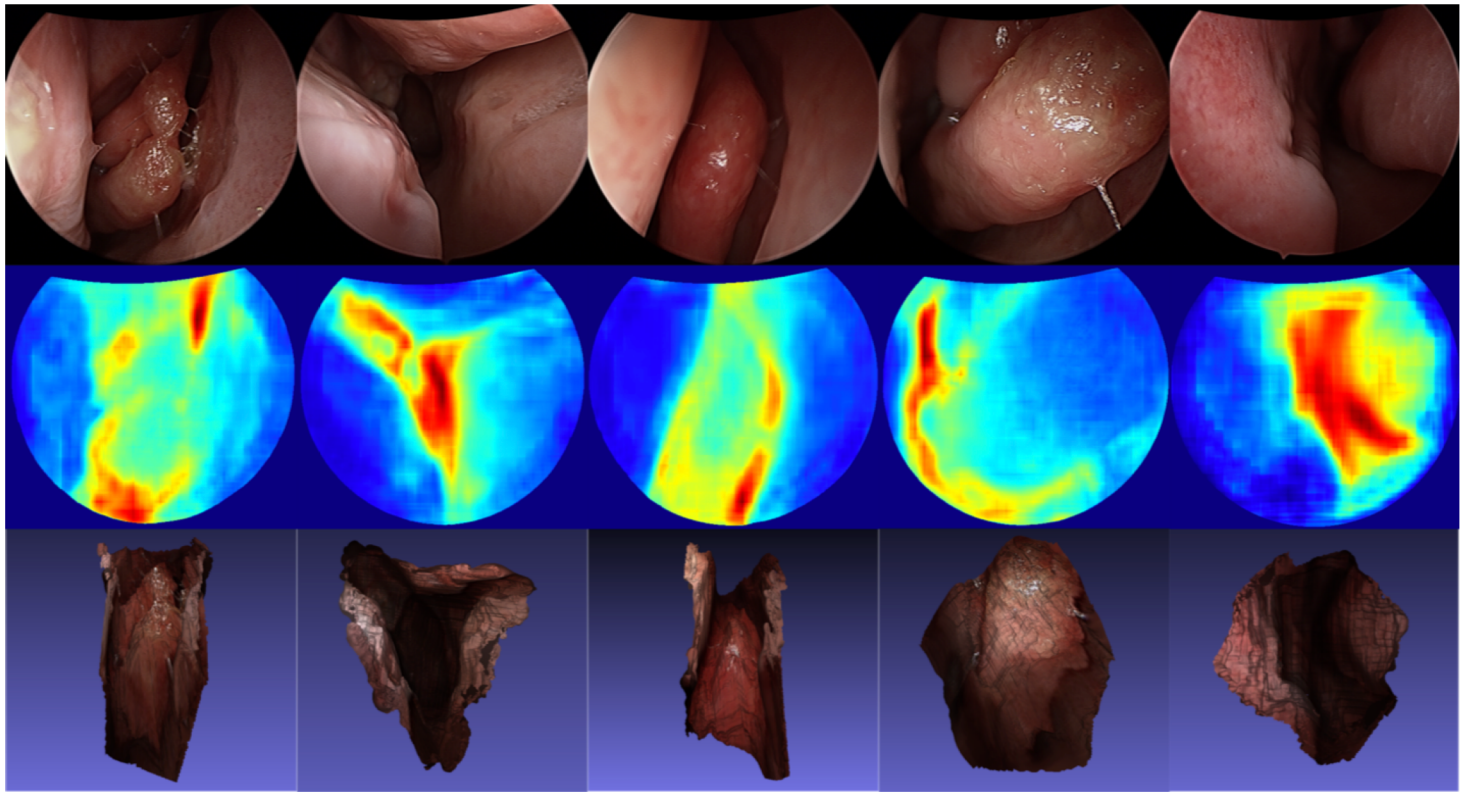
JOHNS HOPKINS
MALONE CENTER *for*
ENGINEERING *in* HEALTHCARE

LABORATORY FOR
Computational
Sensing + Robotics

Dataset Collection

Prior Work: 3D reconstruction from endoscopic surgical video

- In short:
  - Take sequence of 2D color images
  - Do Structure-from-Motion
  - Infer depth from single image
- Accurate to 0.53mm-1.12mm



[3] Reiter, Austin, et al. "Endoscopic-CT: learning-based photometric reconstruction for endoscopic sinus surgery." Medical Imaging 2016: Image Processing. Vol. 9784. International Society for Optics and Photonics, 2016.
[4] arXiv:1806.09521 [cs.CV]

JOHNS HOPKINS
MALONE CENTER *for*
ENGINEERING *in* HEALTHCARE

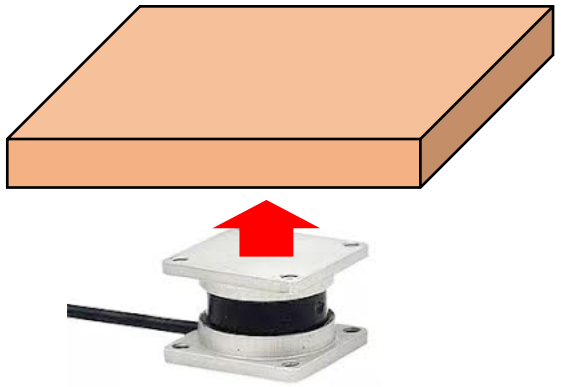
LABORATORY FOR
Computational
Sensing + Robotics

Dataset Collection

Force Data Collection:

- OptoForce 3D Force Sensor

Place underneath the object



- Accuracy: $12.5\times10^{-3}\ N$
- Observation: z-component only
- Re-bias each time we place an object

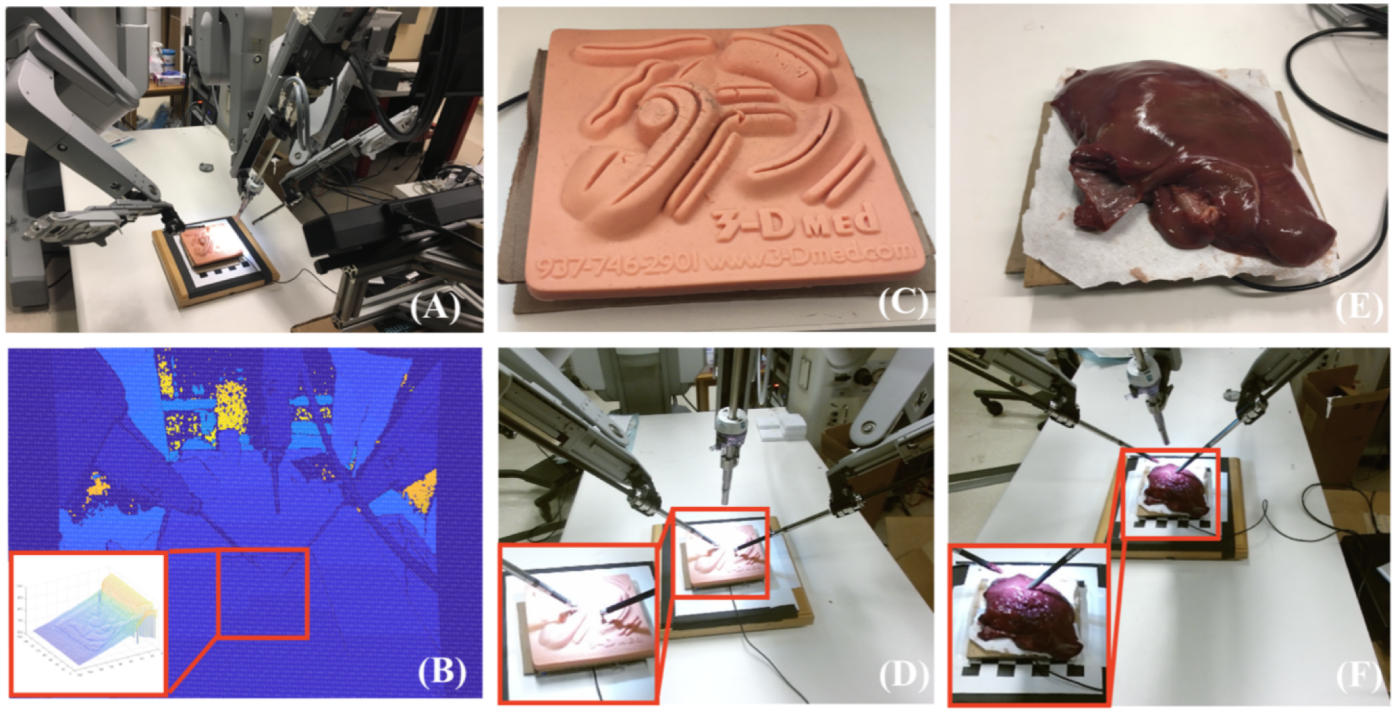


Fig. 2. (A) The experimental environment. (B) The depth image from the Kinect2 camera. (C) The phantom used in the experiment. The force sensor is placed under the phantom. (D) RGB image of the phantom. (E) A fresh piece of pig liver. (F) RGB image of the liver.

JOHNS HOPKINS
MALONE CENTER *for*
ENGINEERING *in* HEALTHCARE

LABORATORY FOR
Computational
Sensing + Robotics

Architecture of RPC-TCN

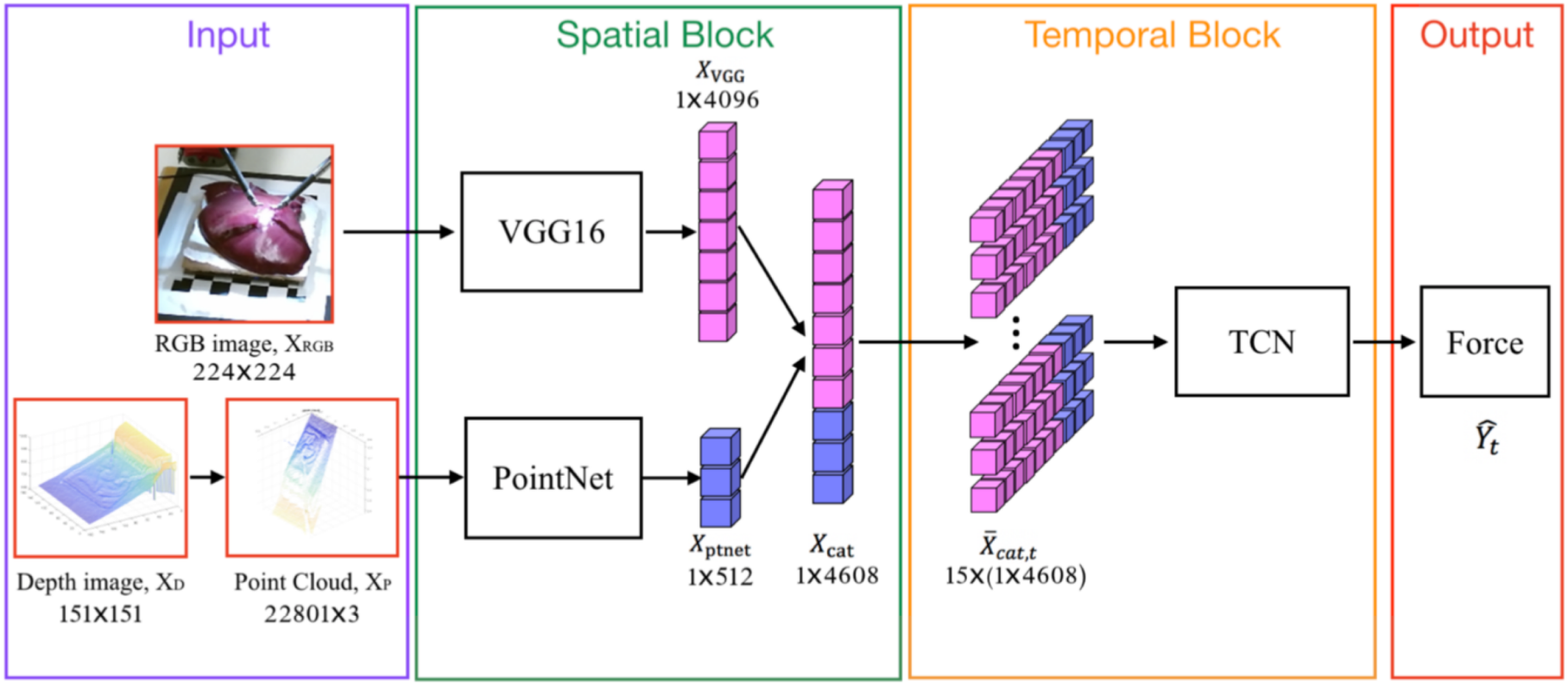


Fig. 3. The basic architecture of RPC-TCN. The spatial block extracts features from the pre-trained VGG net and PointNet and concatenates two features. The temporal block expands this feature to be 15 frames in a window and predict the force corresponding to the middle frame.

JOHNS HOPKINS
MALONE CENTER *for*
ENGINEERING *in* HEALTHCARE

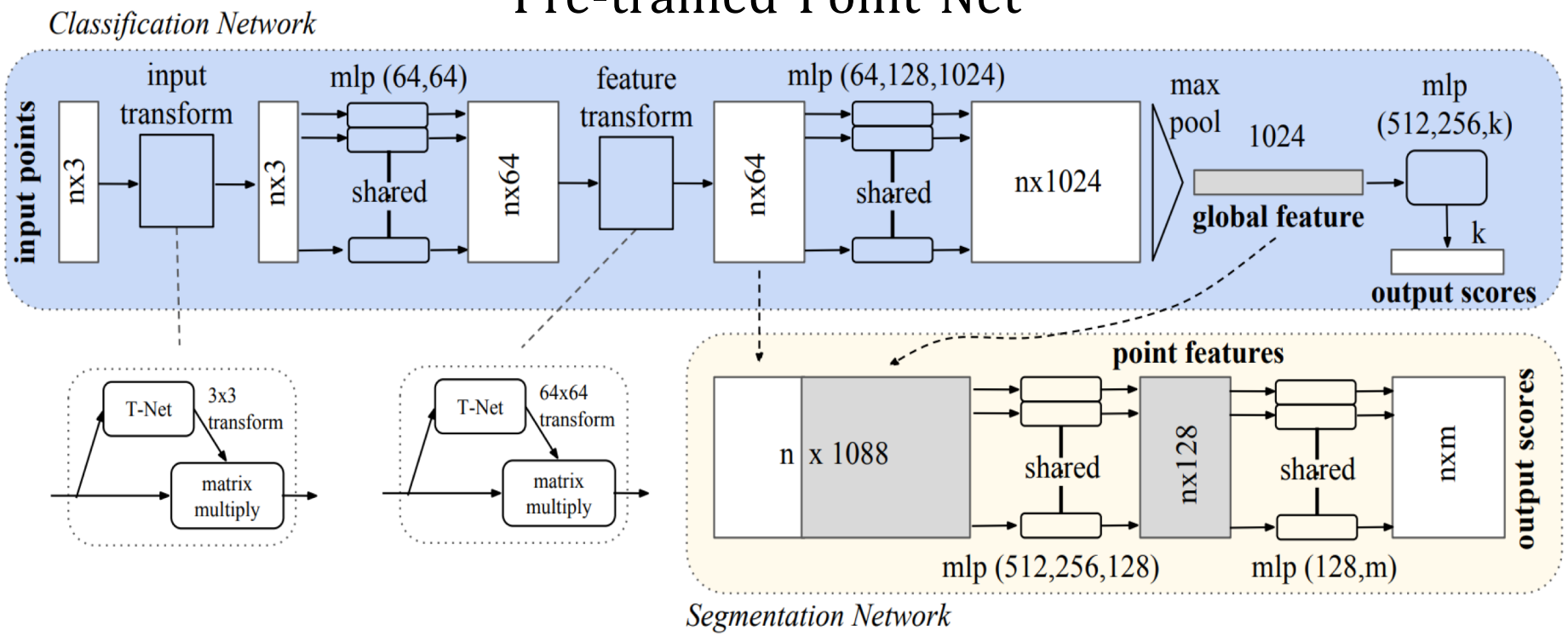
LABORATORY FOR
Computational
Sensing + Robotics

Depth Image to Point Cloud

Depth Image -> 3D Point Cloud

- Invariant to Camera View Points

Pre-trained Point Net



[5] Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." Proc. Computer Vision and Pattern Recognition (CVPR), IEEE 1.2 (2017): 4.

Temporal Block

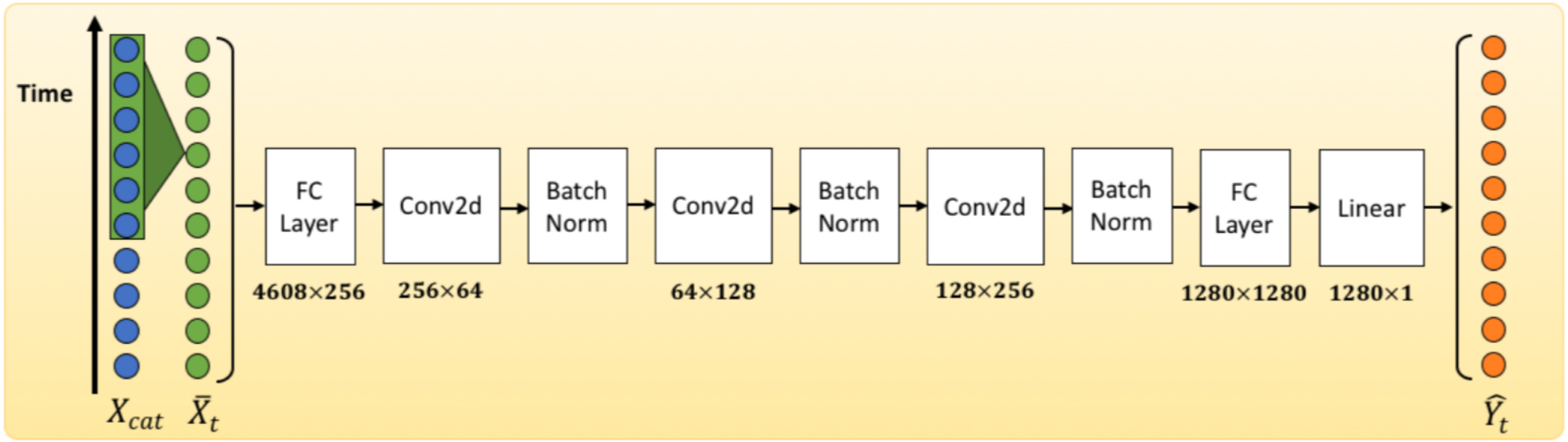


Fig. 4. Hierarchical structure of the temporal block.

Experiments & Results

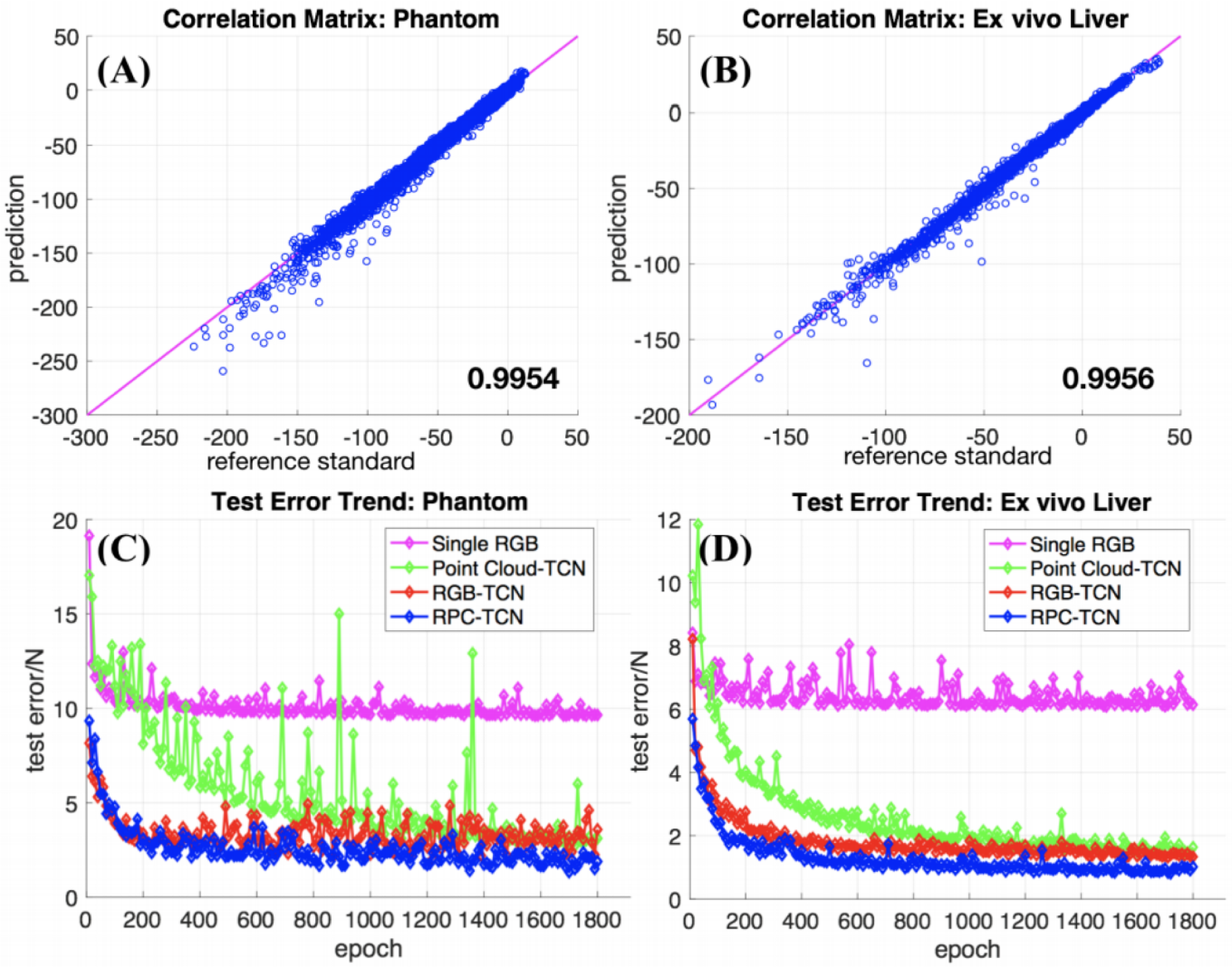


Fig. 5. (A) (B) Illustration of the correlation matrix between reference standard force and the prediction force. (C) (D) Test Error trend with training epochs. The error is calculated as mean absolute error of all test data.

Phantom study: 61,473 samples

Ex vivo study: 44,413 samples

80% for training

5% for validation

15% for testing

Loss Function: Mean Square Error

JOHNS HOPKINS
MALONE CENTER *for*
ENGINEERING *in* HEALTHCARE

LABORATORY FOR
Computational
Sensing + Robotics

Experiments & Results

Table 1. Ablation study results.

Algorithm	Mean Absolute Error (N)		Percentage Error	
	Phantom	*Ex vivo* Liver	Phantom	*Ex vivo* Liver
Single-frame RGB	7.06	10.4	3.01%	5.45%
RGB-TCN	2.51	1.74	1.05%	0.913%
Point Cloud-TCN	2.14	1.87	0.896%	0.983%
RPC-TCN	**1.45**	**0.814**	**0.604%**	**0.427%**

The percentage error is based on the maximum force magnitude, which is -239N for the phantom study and -190N for the *ex vivo* liver study.

JOHNS HOPKINS
MALONE CENTER *for*
ENGINEERING *in* HEALTHCARE

LABORATORY FOR
Computational
Sensing + Robotics

Experiments & Results

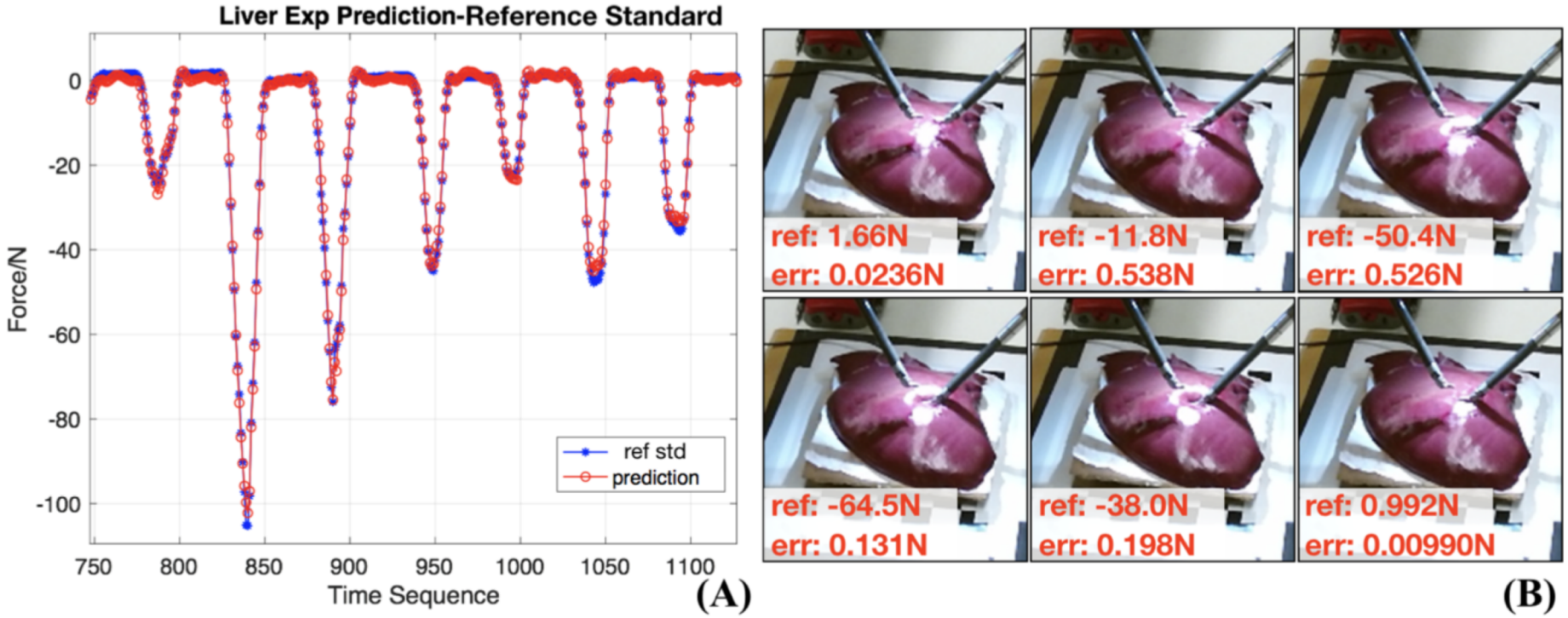


Fig. 1. Results of the *ex vivo* liver study: (A) Test result w.r.t time sequence. The blue line represents the reference standard force and the red curve are the model predictions. (B) A set of screenshots of the RGB image and their reference standard force and error.

JOHNS HOPKINS
MALONE CENTER *for*
ENGINEERING *in* HEALTHCARE

LABORATORY FOR
Computational
Sensing + Robotics

Discussion

- Large force magnitudes are intentionally tested to predict the onset of excessively large force and, warn clinicians.

- Training and testing images include specularities, which improves the generalization ability to different surgical scenarios.

- Our current model does not evaluate the transferability to different organs. Current results are reached by training and testing on one single phantom and *ex vivo* liver.

Future Work

- Future study will include testing on multiple organs and considering tissue biomechanical properties.

- We will consider setups that are more realistic regarding clinical practice, including monocular depth estimation from RGB endoscopic video and slave-side force sensors to accurately measure tool tip contact force.