Deep Learning-Based Image Segmentation on Multimodal Medical Imaging

Presenter: Huidong Xie.

Contribution

- CNNs have been applied to segmentation of tumors in brain, liver, breast, lung, etc.
- There has been little investigation from a systematic perspective about how multimodal imaging should be used. They address this problem by testing different fusion strategies through different CNNs.

Dataset

- They use the publicly available soft-tissue sarcoma (STS) dataset. The dataset contains images from 3 imaging modalities: PET, CT and MRI (both T1-weighted and T2-weighted). They treat T1 and T2 MRI images as 2 different modalities since they portray different tissue characteristics.
- 28*28 patches are extracted for training.
- A patch is labeled as “positive” if its center pixel is within tumor and negative otherwise.
- They randomly selected negative patches to the same number of positive patches.

Example of multimodal images

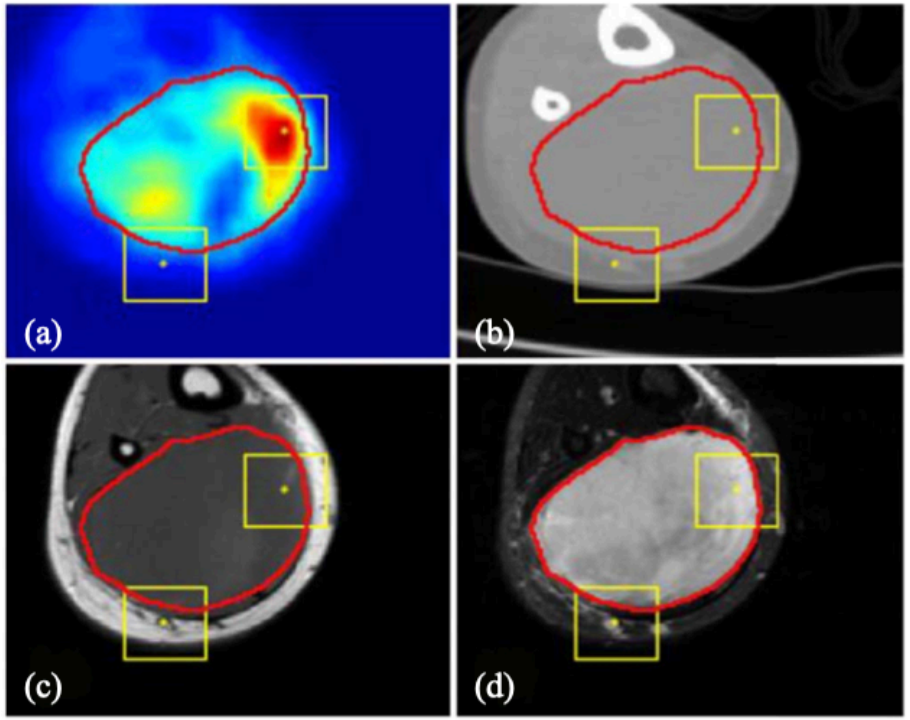

Fig. 1. Multimodal images on the same position from a randomly selected subject. (a) PET; (b) CT; (c) T1; and (d) T2. The image size of this subject is 133×148. Red line is the contour of ground truth from manual annotation. Two yellow boxes illustrate the size of patches (28×28) used as the input for CNN. The center pixel of one patch is within the tumor region and another patch outside the tumor region.

Different fusing strategies

- Fusing at feature level: multimodality images are used together to learn a unified image feature set.
- Fusing at the classifier level: images of each modality are used as separate inputs to learn individual feature sets.
- Fusing at the decision-making level: images of each modality are used independently to learn a single-modality classifier. Final decision is obtained by fusing the output from all the classifiers.

3 types of CNNs

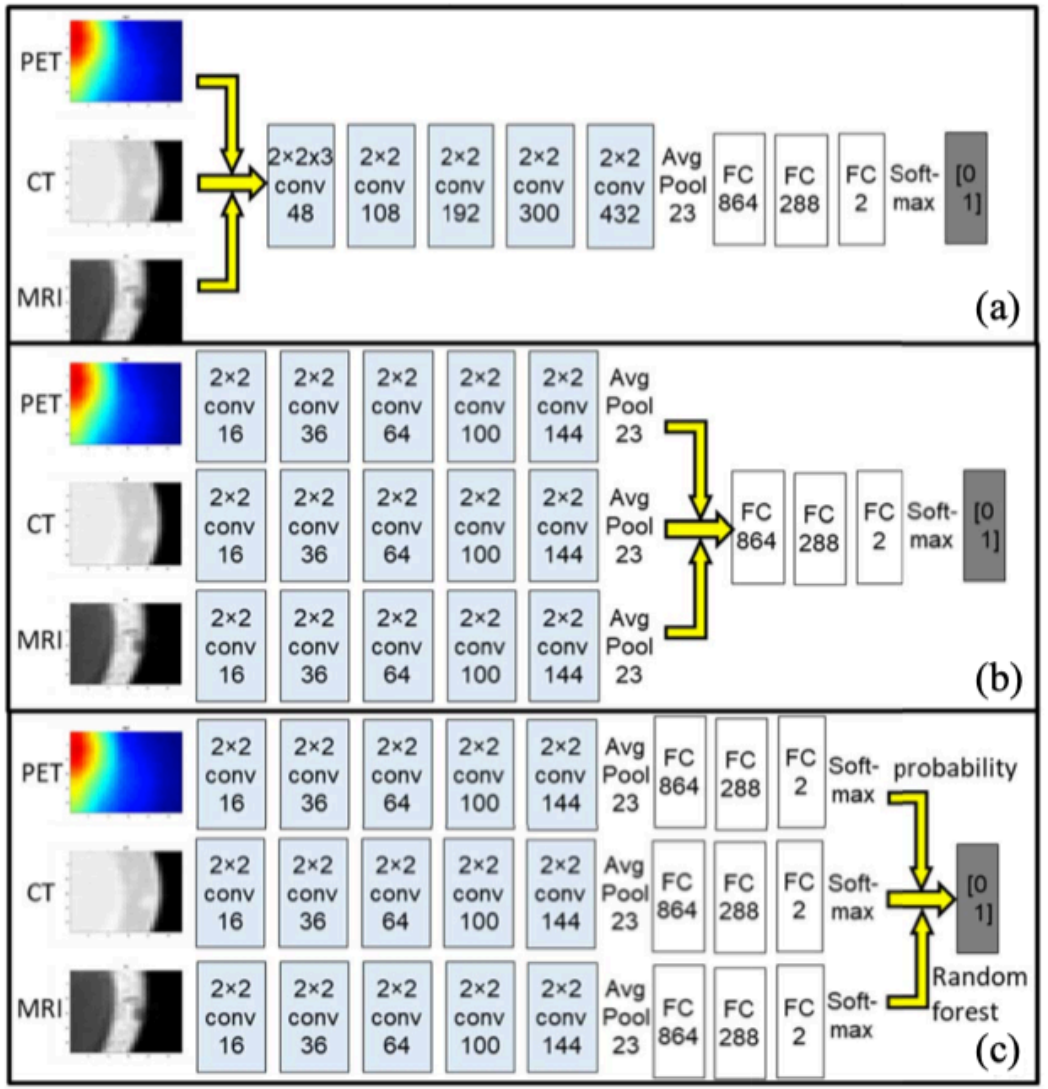

Fig. 2. Illustration of the structure for (a) Type-I fusion networks, (b) Type-II fusion network and (c) Type-III fusion network. The yellow arrows indicate the fusion location.

Results (between single modality and multi-modality)

Sørensen–Dice coefficient (DICE coefficient) which equals to twice the number of voxels within both regions divided by the summed number of voxels in each region to measure the similarity between predicted region and annotation region.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

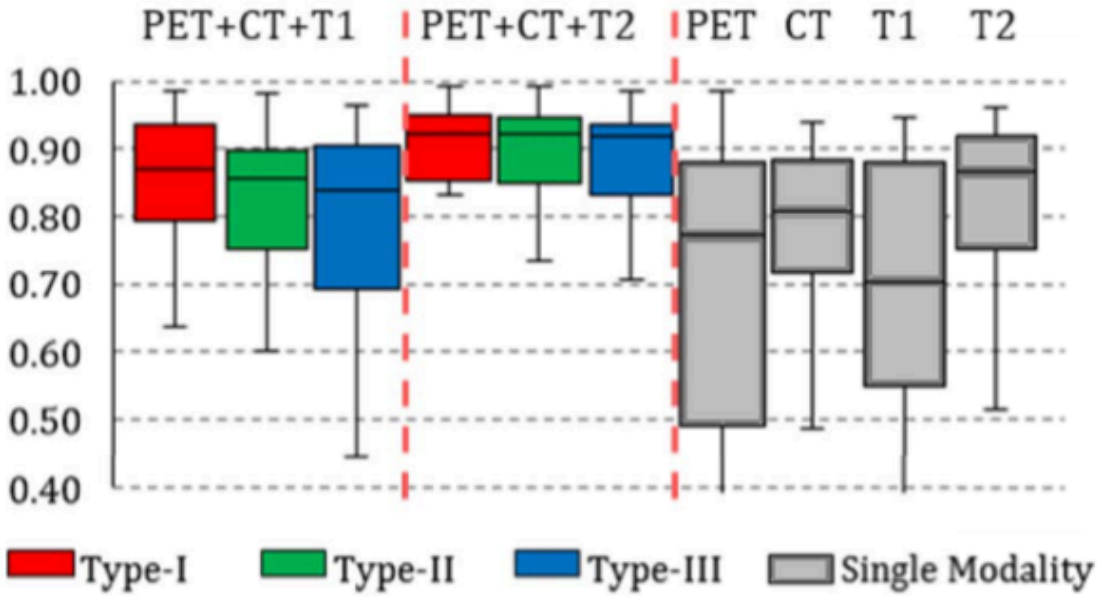

Fig. 4. Box chart for the statistics (median, first/third quartile and the min/max) of the DICE coefficient across 50 subjects. Each box corresponds to one specific type of network trained and tested on one specific combination of modalities. For example, the first box from the left shows the prediction statistics of Type-I fusion network trained and tested on images from PET, CT, and T1-weighted MR imaging modalities.

Results (for synthetic low-quality image)

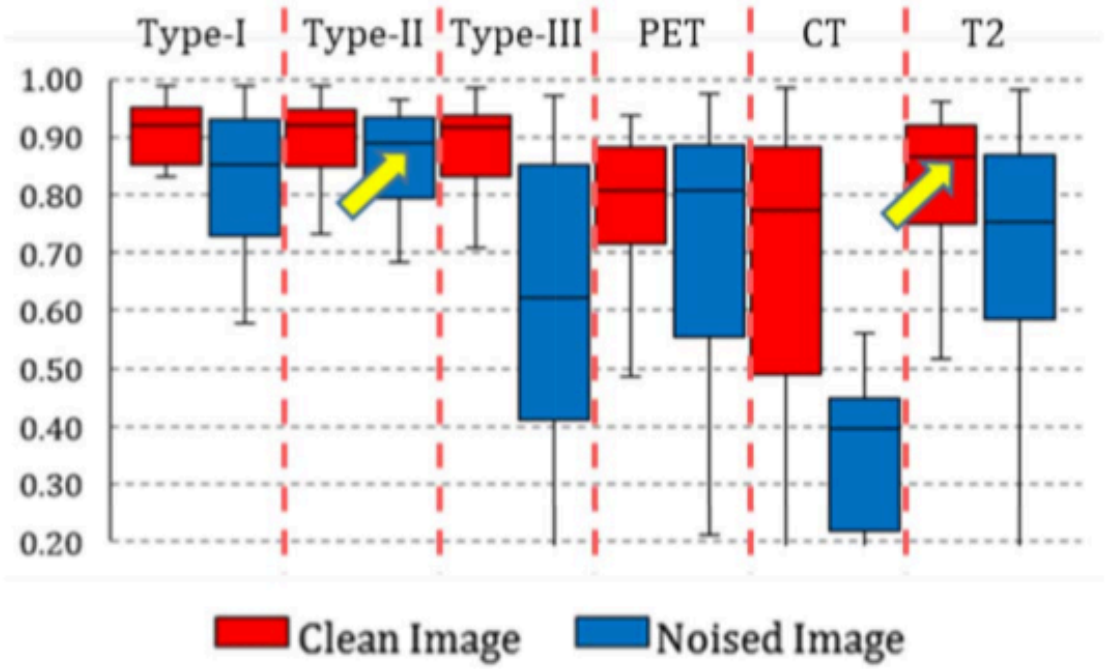

Fig. 6. Box chart for the statistics (median, first/third quartile and the min/max) of the DICE coefficient across 50 subjects. Red box stands for networks trained and tested on original clean images and blue box stands for networks based on synthetic noised image.

Results (for different modality combinations)

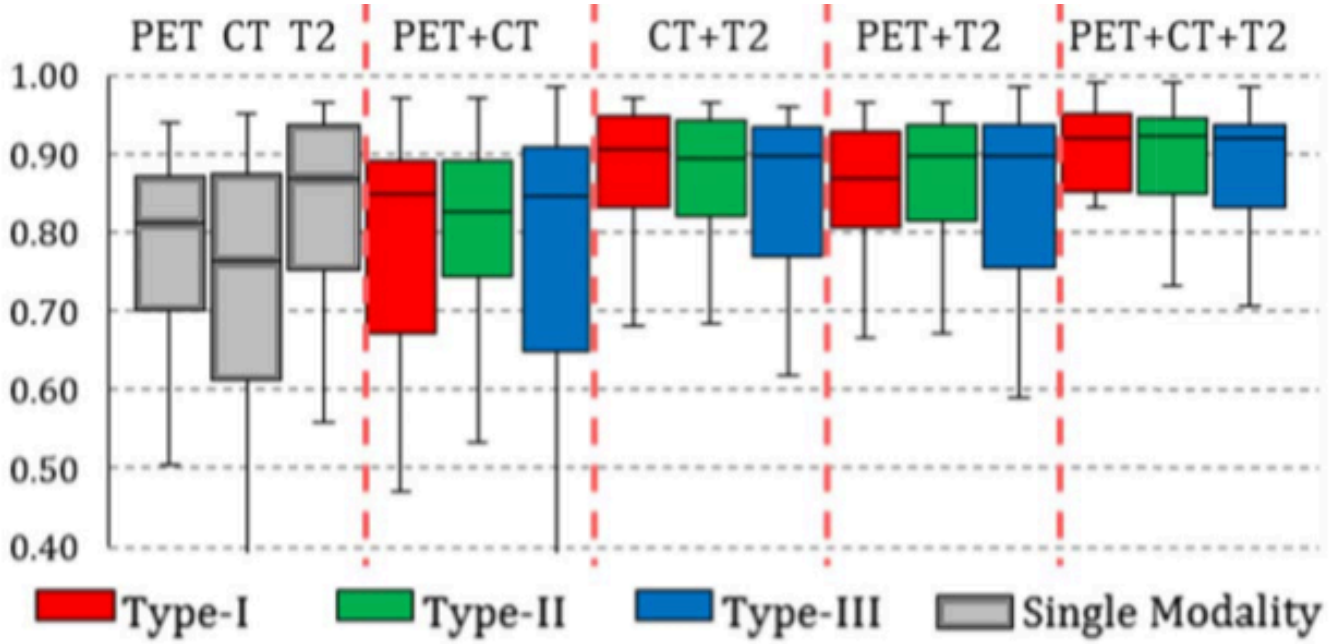

Fig. 8. Box chart for the statistics (median, first/third quartile and the min/max) of the DICE coefficient across 50 subjects. Red box stands for network train and test on Type-I network, blue box stands for Type-II network, and green stands for Type-III. Performances of single-modality network are shown as gray boxes to the left for reference.