A semi-automated method to separate tumor from physiological uptakes on FDG PET-CT for efficient generation of training data targeting deep learning

Kenji Hirata[1], Sho Furuya[1], Sung-Cheng Huang[2], Osamu Manabe[1], Keiichi Magota[3], Kentaro Kobayashi[1], Shiro Watanabe[1], Takuya Toyonaga[4], Nagara Tamaki[5], Tohru Shiga[1]

[1]Department of Nuclear Medicine, Hokkaido University, Sapporo, Japan, [2]Department of Molecular and Medical Pharmacology, UCLA, Los Angeles, CA, [3]Division of Medical Imaging and Technology, Hokkaido University Hospital, Sapporo, Japan, [4]Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, [5]Department of Radiology, Kyoto Prefectural University of Medicine, Kyoto, Japan

Deep learning needs training data

Segmentation of malignant lesions on FDG PET-CT is one of the most expected tasks for deep learning. To develop an efficient deep-learning-based system, a massive database of training data, usually defined by radiologists, is of great importance. Compared to CT and MRI, definition of tumor boundary on PET is easier as various methods of automated segmentation, such as fixed/adaptive threshold and gradient methods, can work.

What is the problem?

Time-consuming manual operation usually leads to low reproducibility for tumors adjacent to physiological or inflammatory uptake.
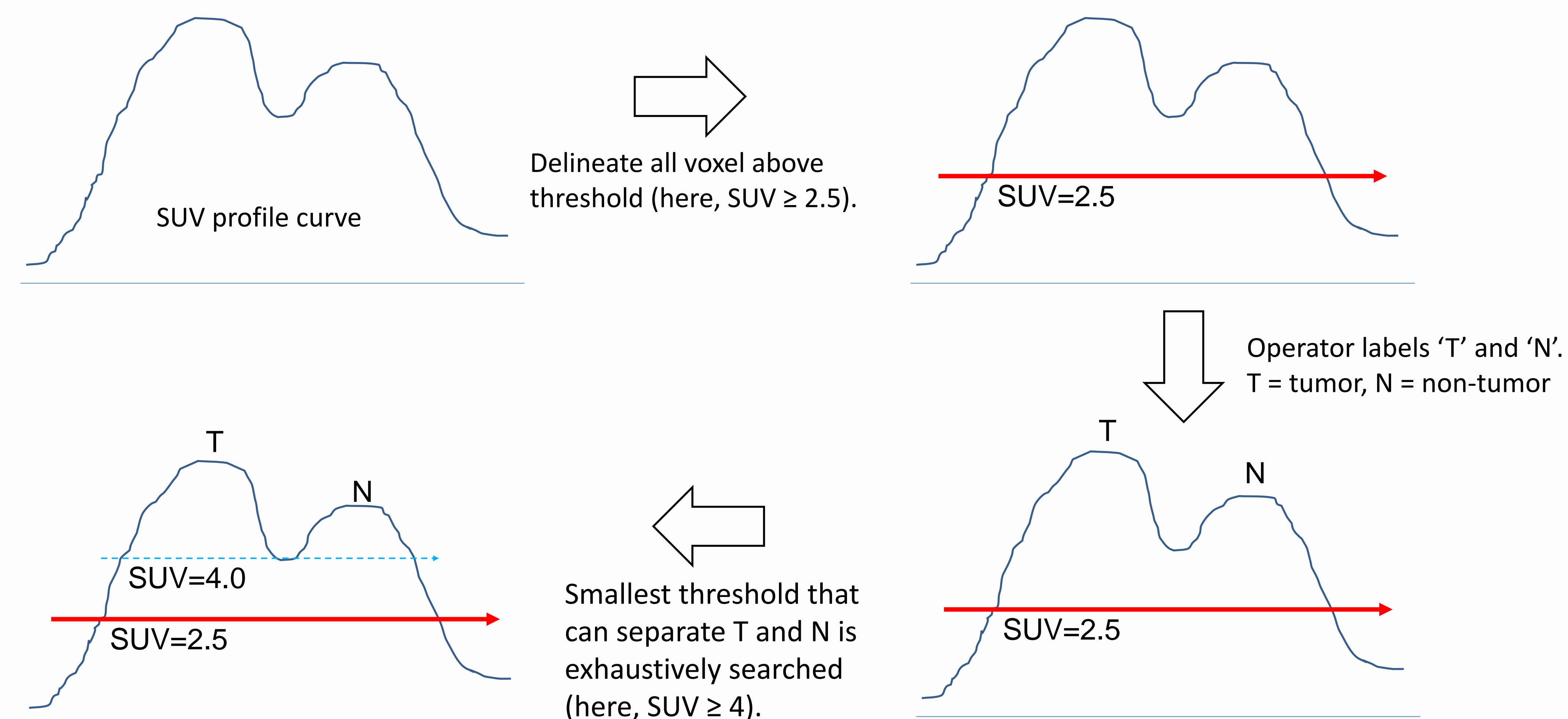
Segmentation
SUV ≥ 2.5

Primary tumor

Contralateral physiological uptake

Metastatic lymph node
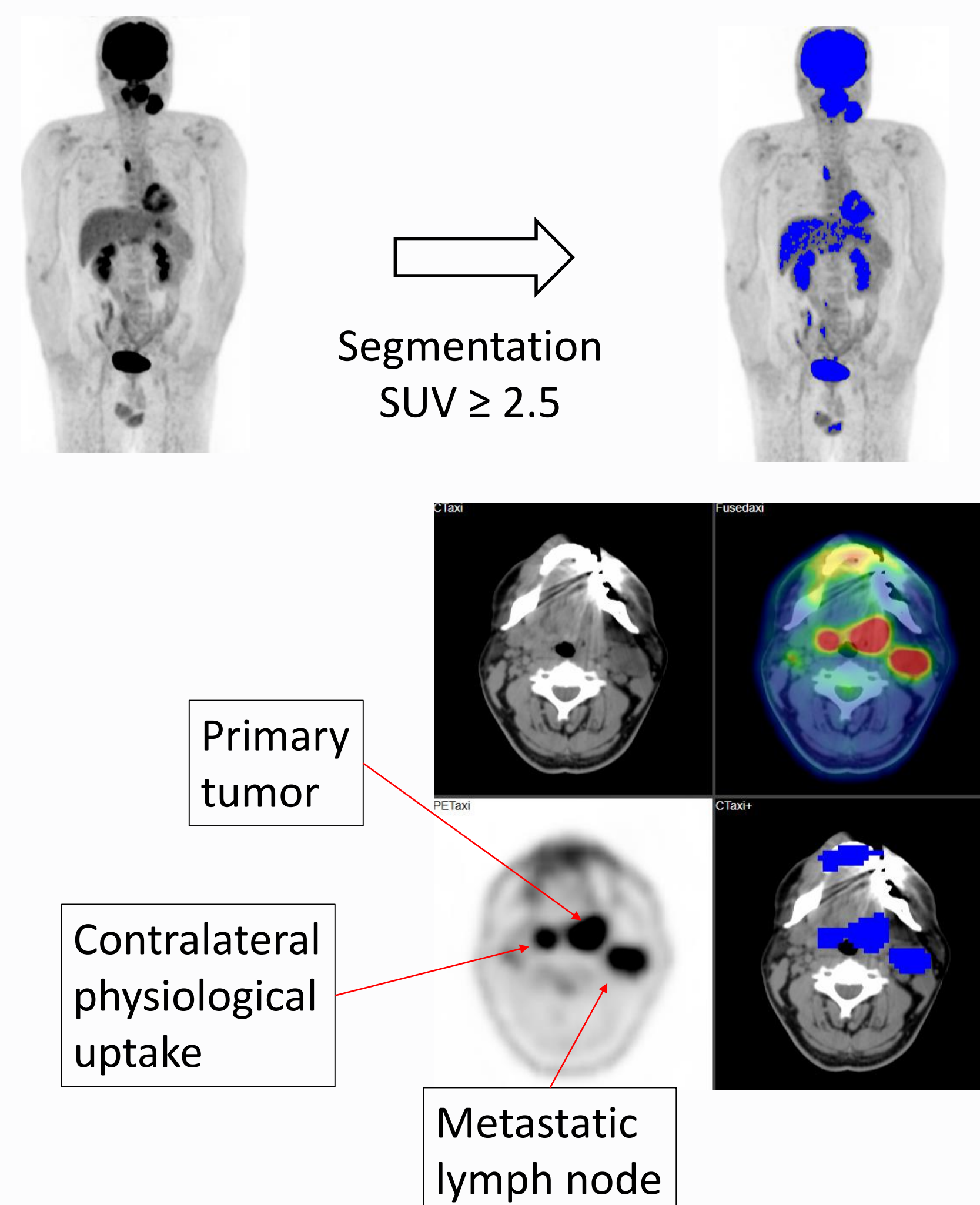
In this case of left oropharyngeal cancer, using threshold of SUV ≥ 2.5, the primary tumor is segmented together with contralateral physiological uptake. Usually, manual interaction is required to solve the problem.

Purpose

We aimed to propose a new method that requires minimal human interaction to delineate tumor uptakes, and to test its inter-operator reproducibility.

Algorithm (one-dimensional illustration)

SUV profile curve

Delineate all voxel above threshold (here, SUV ≥ 2.5).

SUV=2.5

Operator labels 'T' and 'N'. T = tumor, N = non-tumor

T   N

SUV=2.5

Smallest threshold that can separate T and N is exhaustively searched (here, SUV ≥ 4).

T   N

SUV=4.0
SUV=2.5

Once an optimum threshold is found, the voxels that are greater than the second threshold (here, 4.0) are automatically labeled with either T or N.

On the other hand, the voxels that are between the first (here, 2.5) and the second (here, 4.0) thresholds need some algorithm before being labeled.

We used a **steepest uphill method** to solve the problem. This is a simple algorithm where a voxel is labeled with the same label of greatest neighbor. For example, let us consider the following voxel array.

1 3 5 7 4 3 2 5 6 9 8

Looking at voxel 2, the greatest neighbor (=connecting) voxel is 5. Similarly, 5 finds 6, and 6 finds 9, and 9 is the local maximum. Therefore, 2 and 9 are labeled with the same label.

In case that no optimum threshold was found, a **nearest neighbor algorithm** was used to solve the label collision problem, where an un-labeled voxel is labeled with the same label of the nearest labeled voxel.

Performance test

Subjects
FDG PET-CT images of patients with head-and-neck cancer (N=13) and gynecological cancer (N=10) were investigated retrospectively, because these cancers frequently co-exist with physiological accumulation.

Segmentation
Two experienced nuclear medicine physicians independently used the current method to segment the primary tumor. The physicians were requested to remove lymph nodes and physiological uptakes. The physicians repeatedly used the current method until the segmentation result looked satisfactory.

Reproducibility evaluation
To evaluate inter-operator reproducibility, dice similarity coefficient (DSC) was employed to compare the 2 volumes of interest (VOI) of each tumor, calculated by DSC = $2 |VOI_1 \cap VOI_2| / (|VOI_1| + |VOI_2|)$, where VOI_1 and VOI_2 were VOI defined by the two physicians, respectively. $|X|$ indicates the number of voxels within the region X. DSC ranges 0 to 1, and a higher value represents a higher similarity of the two VOIs (i.e., better reproducibility).
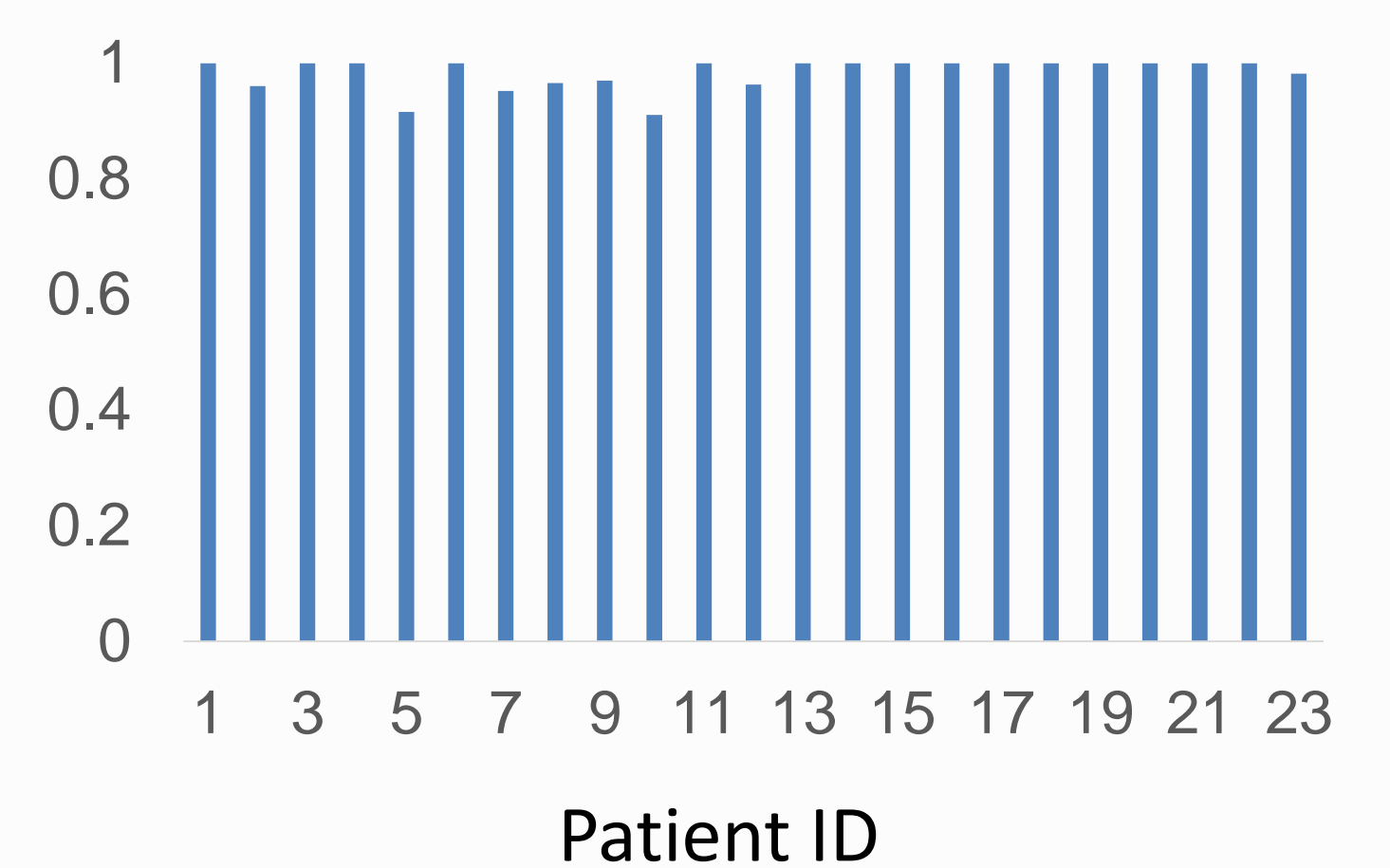
Results

The method worked successfully for all the patient images.
Metabolic tumor volume (MTV) was 96 ± 127 mL vs. 97 ± 126 mL (Physicians 1 vs. 2, mean ± SD) with Pearson's correlation coefficient R being 0.9968. Similarly, total lesion glycolysis (TLG) was 580 ± 720 mL vs. 584 ± 723 mL, respectively (R = 0.9986).
DSC was 0.98 ± 0.03, ranging from 0.91 to 1.00, indicating very high similarity of the VOIs. Of 23 patients, DSC was 1.0 (i.e., perfect match) in 14 patients (6/13 head-and-neck cancer and 8/10 gynecological cancer).
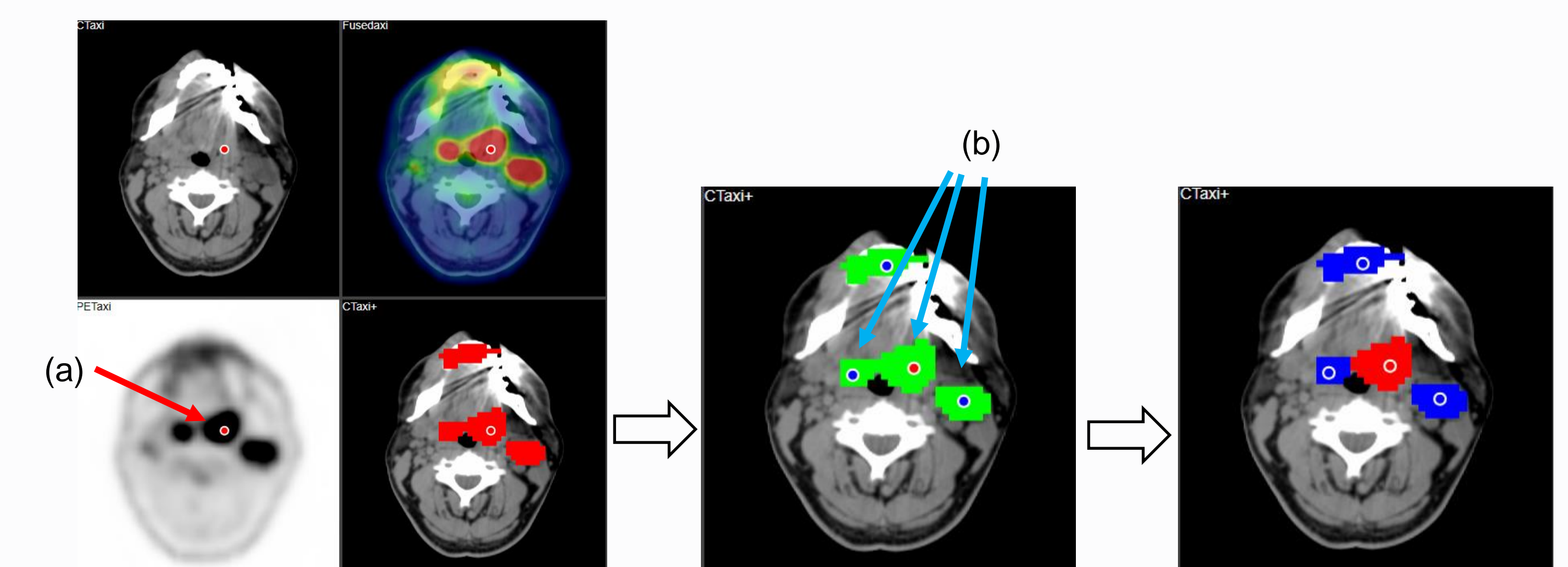
Metabolic tumor volume (MTV)

y = 1.0003x
R² = 0.9968

Physician 2 / Physician 1

Dice similarity coefficient (DSC)

Patient ID
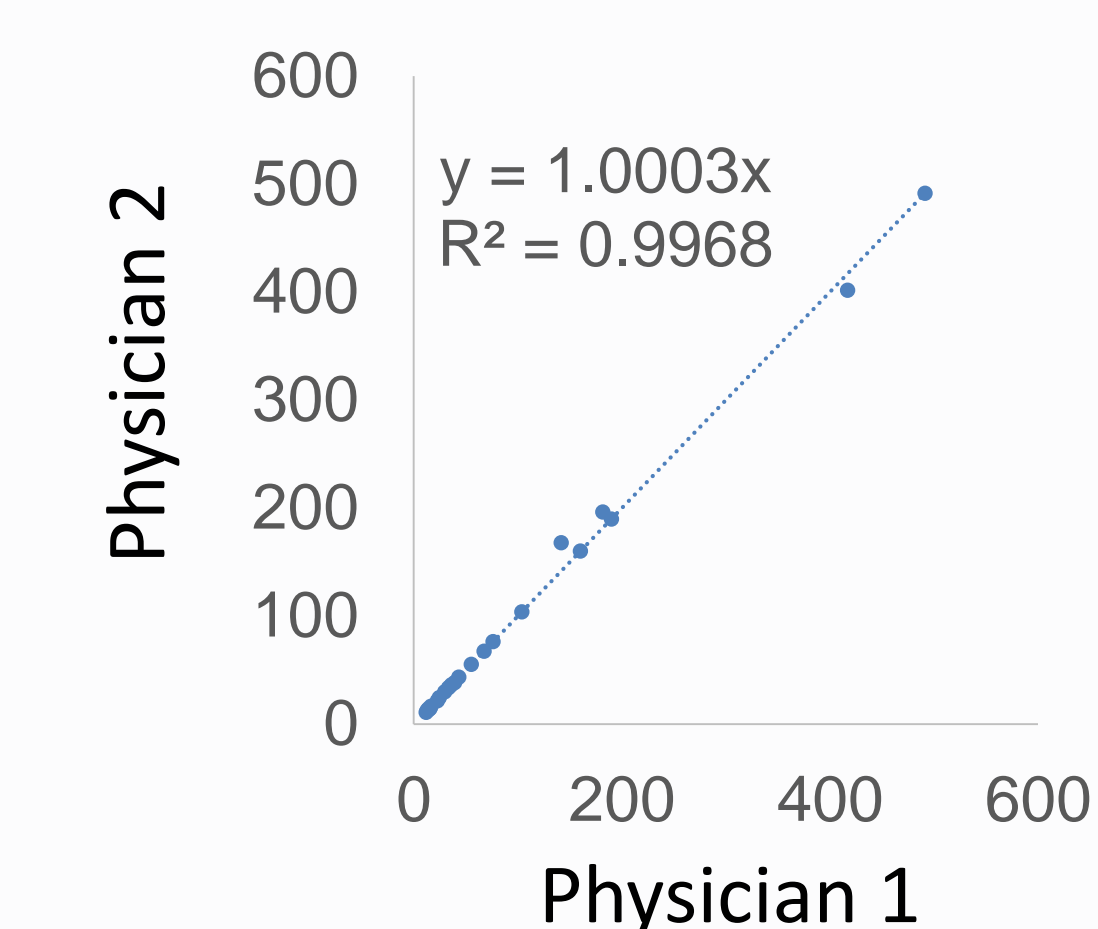
Representative case

(b)

(a)

The operator placed a red mark in the primary tumor (a), resulting both physiological uptake and metastatic node (connected on other slices) turning red. Then, blue marks were placed in physiological and nodal uptake (b, turning green), and solved the label collision problem.

Conclusion

Using our new method, physicians successfully segmented the tumors from physiological or inflammatory uptakes with high inter-operator reproducibility.

This semi-automated method will not only help prepare training data for FDG PET-CT, but will contribute to MTV, TLG, and radiomic analysis.