Deep Image-to-Image Translation for Spatial Alignment in Sequential Dual-Energy CT

Maximilian Hammermann[1,2], Elias Eulig[1,2], Joscha Maier[1], and Marc Kachelrieß[1,2]

[1]Division of X-Ray Imaging and CT, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2]Heidelberg University, Heidelberg, Germany

Purpose

- Dual-energy computed tomography (DECT) is an important diagnostic modality, *e.g.* due to its improved material differentiation capabilities by acquiring two CT scans at different detected spectra.
- State-of-the-art DECT systems (*e.g.* dual-source or fast tube voltage switching CT) are rather expensive and therefore not widely available in clinical environments.
- Sequential dual-energy measurements may be acquired on every CT scanner, but the resulting image data may be misregistered due to the occurence of patient motion and not even be fully aligned after applying conventional deformable image registration algorithms.
- Goal of this work: Investigate the use of deep image-to-image translation methods to derive aligned dual-energy information from coarsely registered image pairs.

Materials and Methods

A) Proposed Approaches

- Investigate deep learning-based strategies, that realize a mapping $\mathcal{M}$ towards aligned DECT image pairs (f_L, f_H):

$$\mathcal{M} : (f_\mathrm{L}, f'_\mathrm{H}) \rightarrow f_\mathrm{H}. \tag{1}$$

- Considered two residual U-net pipelines processing low-energy scans f_L as well as high-energy scans with misalignments f'_H as a dual-channel input.
- USpectNet directly predicts aligned high-energy images.
- W-USpectNet estimates parameters for a weighted linear transformation, that is applied on low-energy images f_L by a custom mapping layer (ML).

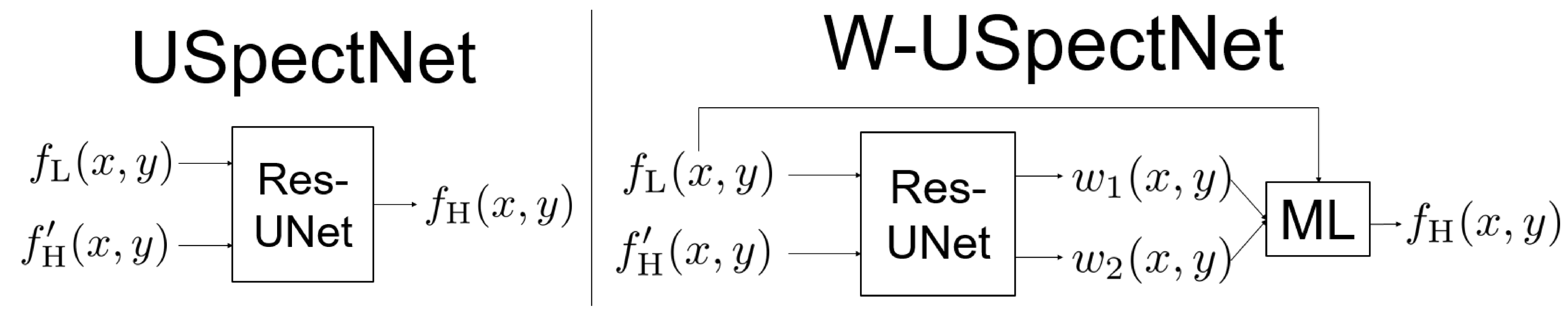

Fig. 1: Schematic overviews of the proposed approaches.

B) Comparison Approaches

- State-of-the-art deformable registration algorithm: SyN[1] (symmetric normalization) from the ANTs toolkit with a mutual information similarity metric.
- Deep learning-based single-energy mapping approach (VirtualDE) to map low- to high-energy images directly.

C) Data Generation, Training and Testing

- Data baseline: 42 patient scans from DECT examinations on a Siemens Somatom Force (70 kV / 150 kV Sn).
- Simulate coarsely registered sequential DECT scans by applying slice-wise random deformation vector fields (maximum displacements of 6 mm per dimension) to one scan of dual-source DECT image pairs.
- Supervised network training using an $\mathcal{L}_1$-loss function.
- Apply a data-augmentation strategy to inhibit the network from learning shortcuts towards common spectral relationships of the different-energy scans.
- Simulate test data in the same way as training data using CT data of eight new patients not used during training.

Results

A) Quantitative Evaluation

- Calculation of standard image quality metrics between the ground truth (GT) and the network output: The mean absolute error (MAE) over patient regions in both images, the structural similarity index measure (SSIM) in a typical diagnostic window for the respective anatomical context and the visual information fidelity measure (VIF).
- Metricwise, our approaches outperform VirtualDE as well as the considered deformable registration algorithm.

Approach	MAE [HU] (↓)	SSIM (↑)	VIF (↑)
Uncorrected	82.9 ± 18.6	0.680 ± 0.108	0.083 ± 0.021
VirtualDE	23.8 ± 1.9	0.810 ± 0.108	0.370 ± 0.036
SyN	22.3 ± 2.9	0.850 ± 0.095	0.331 ± 0.050
USpectNet	11.3 ± 1.6	0.960 ± 0.033	0.526 ± 0.037
W-USpectNet	10.9 ± 1.7	0.963 ± 0.031	0.535 ± 0.038

Tab. 1: Mean and standard deviations of image quality metrics over the entire test set.

B) Qualitative Comparison

- The upper image row shows an example, where the SyN-algorithm fails to align small anatomical structures. Our approaches instead produce well-aligned results.
- The lower image row shows an example, where the VirtualDE-network fails for an oral contrast agent in digestive organs, since the training data of no considered DL-approach contained such samples. In contrast, our approaches make use of the additional spectral input and predict images with correct CT-values.

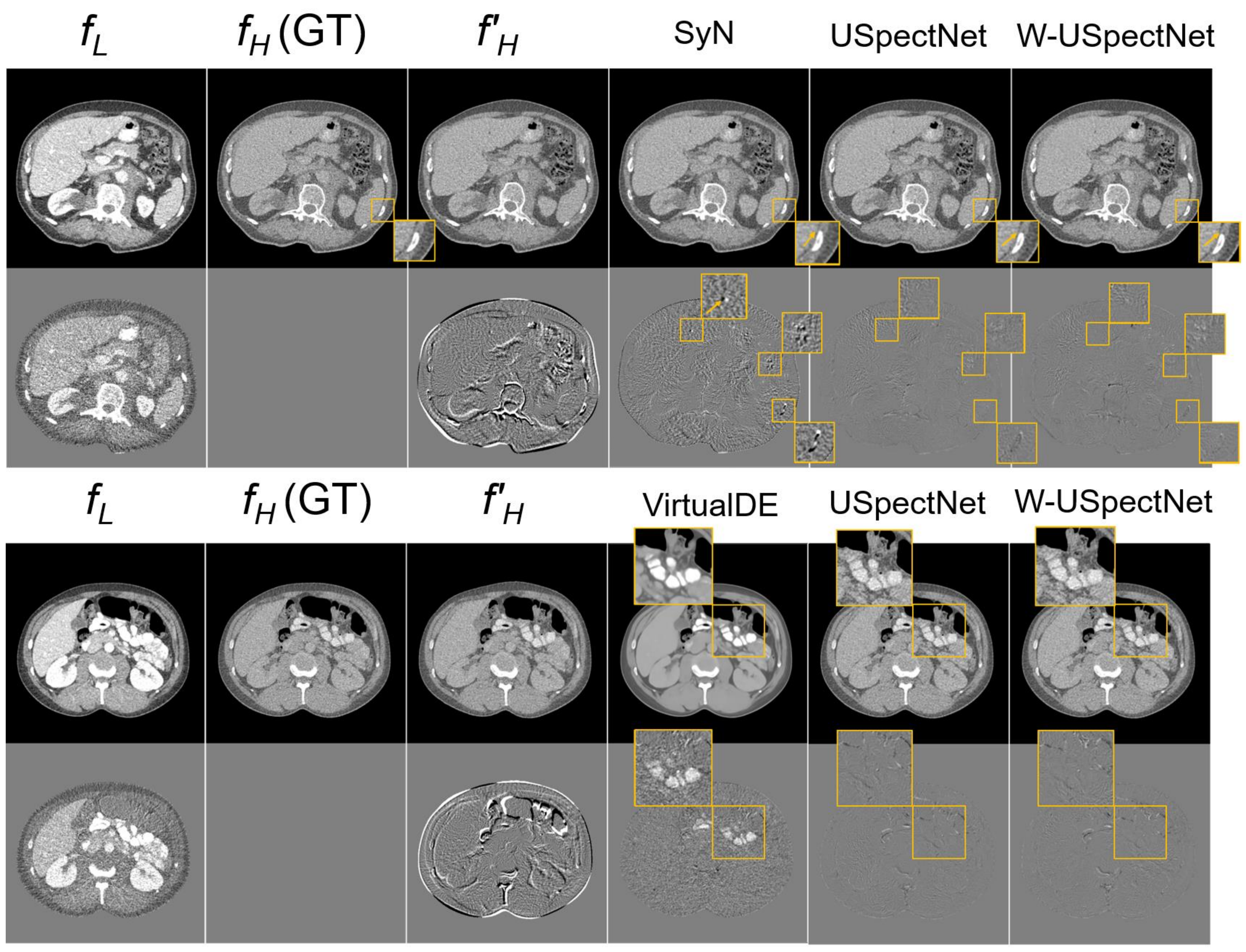

Fig. 2: Images for two patients. The difference images are the difference to the GT (f_H). Images: C = 0 HU, W = 500 HU, Difference images: C = 0 HU, W = 400 HU.

Conclusions

- The proposed networks predict aligned DECT image pairs and outperform the reference methods.
- Adaption of the proposed approaches to 3D as well to the scenario of subtraction imaging is ongoing.

[1]Avants B. B. et al. Medical Image Anal. 2008 Feb; 12(1):26-41.

Correspondence: maximilian.hammermann@dkfz.de