Unsupervised Cross-modal Synthesis of Subject-specific Scans

Raviteja Vemulapalli
Department of Electrical and Computer Engineering
University of Maryland, College Park

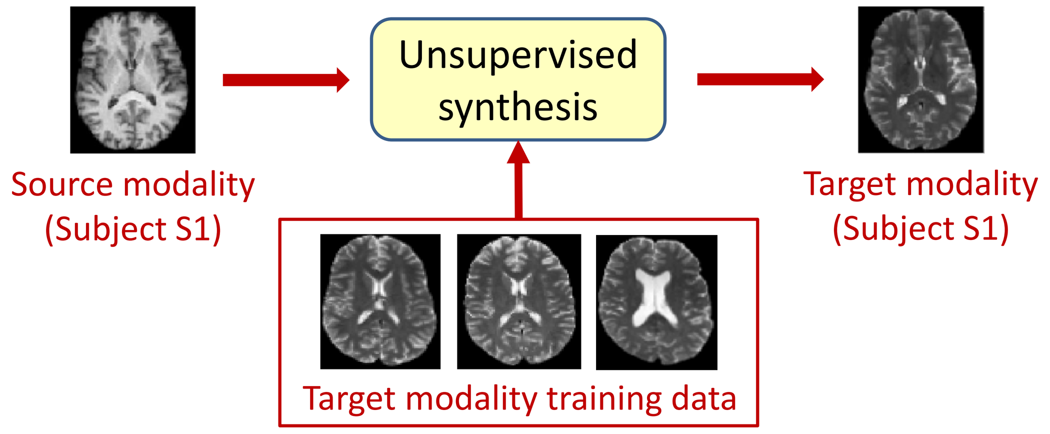
Hien Van Nguyen, Shaohua Kevin Zhou
Siemens Healthcare Technology Center
Princeton, New Jersey

Supervised Cross-modal Medical Image Synthesis

- Existing cross-modal medical image synthesis approaches are supervised, i.e., they require training data from both source and target modalities from the same set of subjects.
- Availability of such paired data is limited in many cases.
- Collecting multiple scans from each subject is undesirable.

Proposed Unsupervised Synthesis Approach

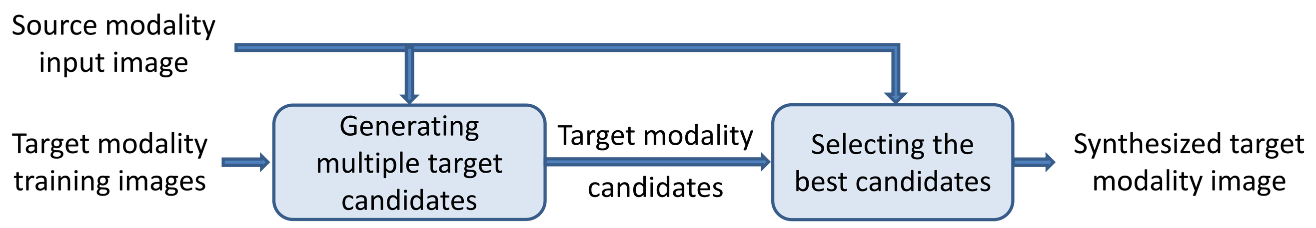
- We propose an unsupervised approach that works without paired training data.



Overview of the Proposed Approach

Two step synthesis strategy

- Generate multiple target modality candidate intensity values for each voxel independently.
- Synthesize a full target modality image by selecting the best candidate values jointly for all the voxels.



Notations

- Φ^v denotes the set consisting of voxel $\boldsymbol{v} = (x, y, z)$ and its neighbors.
- $\Phi^v(p, q, r)$ refers to the voxel $(x+p, y+q, z+r)$.
- X_s and X_t are random variables with support $\psi = \{l_1, \dots, l_L\}$, representing the voxel intensity values of source modality image I_s and target modality image I_t, respectively.

Candidate Generation

- Cross-modal nearest neighbor search for candidate generation:
  - For each voxel $\boldsymbol{v}$, extract a $d_1 \times d_1 \times d_1$ patch centered on $\boldsymbol{v}$ from the given source modality image.
  - Find its K nearest $d_1 \times d_1 \times d_1$ target modality patches by searching across the target modality training images using mutual information as the cross-modal similarity measure.
  - The intensity values of the center voxel and its neighbors from these K nearest patches provide K target modality candidate values for the set Φ^v.

Full Image Synthesis

- Given K target modality candidate values $\{\phi^{v1}, \dots, \phi^{vK}\}$ for the set Φ^v at each voxel,
  - we select one among the K candidates at each voxel, and
  - use the intensity values of the center voxels from the selected candidates to synthesize the target modality image.
- We solve the selection problem jointly for all the voxels based on the following two criteria:
  - **Global mutual information maximization:** The synthesized target modality image should have high mutual information with the given source modality image.
  - **Local spatial consistency maximization:** The candidate selected at each voxel should be spatially consistent with the candidates selected for its neighbors.

Mutual Information Maximization

- Let $w_{vk} = \mathbb{I}[\text{Candidate } \phi^{vk} \text{ is selected at voxel } \boldsymbol{v}]$.

$$\max_{W} \; H(X_t) - H(X_s, X_t)$$

$$H(X_t) = -\sum_{b=1}^{L} P(X_t = l_b) \log[P(X_t = l_b)],$$

$$P(X_t = l_b) = \frac{1}{N} \sum_{v} \sum_{k=1}^{K} w_{vk} \mathbb{I}[\phi^{vk}(0,0,0) = l_b],$$

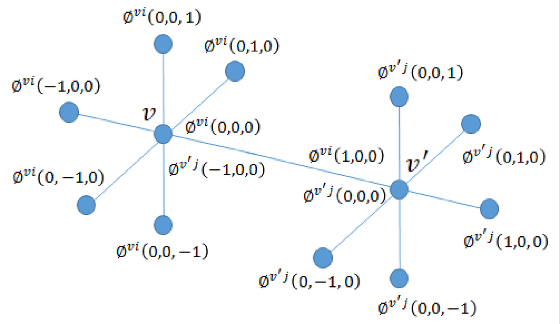
$$H(X_s, X_t) = -\sum_{a,b=1}^{L} P(X_s = l_a, X_t = l_b) \log[P(X_s = l_a, X_t = l_b)],$$

$$P(X_s = l_a, X_t = l_b) = \frac{1}{N} \sum_{v} \sum_{k=1}^{K} w_{vk} \mathbb{I}[I_s(v) = l_a, \phi^{vk}(0,0,0) = l_b].$$

Spatial Consistency maximization

$$\max_{W} \; SC(W)\,;\; SC(W) = -\sum_{vv'} [w_{v1} \dots w_{vK}] \begin{bmatrix} C_{11}^{vv'} & \dots & C_{1K}^{vv'} \\ \vdots & \ddots & \vdots \\ C_{K1}^{vv'} & \dots & C_{KK}^{vv'} \end{bmatrix} \begin{bmatrix} w_{v'1} \\ \vdots \\ w_{v'K} \end{bmatrix},$$

$$C_{ij}^{vv'} = \sqrt{\left(\phi^{vi}(0,0,0) - \phi^{v'j}(v - v')\right)^2 + \left(\phi^{v'j}(0,0,0) - \phi^{vj}(v - v')\right)^2}$$



Combined Formulation

$$\max_{W} \; H(X_t) - H(X_s, X_t) + \lambda SC(W)$$

$$\text{subject to} \quad w_{vk} \in \{0,1\}, \quad \sum_{k=1}^{K} w_{vk} = 1.$$

- We relax the binary constraints on w_{vk} to positivity constraints $w_{vk} \geq 0$, and solve the relaxed optimization problem using reduced gradient ascent method.
- All w_{vk} are initialized with a value of $1/K$ (all candidates are given equal weight at the beginning of the optimization).
- Once we obtain w_{vk}, we use $\phi^{vk^*}(0,0,0)$ to synthesize voxel $\boldsymbol{v}$, where $k^* = \operatorname{argmax}_k w_{vk}$.

Extension to Supervised Setting

- In supervised setting, we have paired training data from source and target modalities.
- Replace the cross-modal nearest neighbor search with source-modal nearest neighbor search:
  - For each voxel $\boldsymbol{v}$, extract a $d_2 \times d_2 \times d_2$ patch centered on $\boldsymbol{v}$ from the given source modality image.
  - Find its K nearest $d_2 \times d_2 \times d_2$ source modality patches by searching across the source modality training images using standard Euclidean distance.
  - Use the target modality patches corresponding to these K source modality neighbors for generating the target modality candidate values for the set Φ^v.

Experiments

- Synthesized T1-MRI scans from T2-MRI scans and vice versa.
- Used coupled sparse representation for post processing.
- Dataset: NAMIC brain database consisting of T1 and T2 MRI scans from 19 subjects.
- Leave-one-out cross-validation setting
- Evaluation metrics: Cross-correlation and PSNR

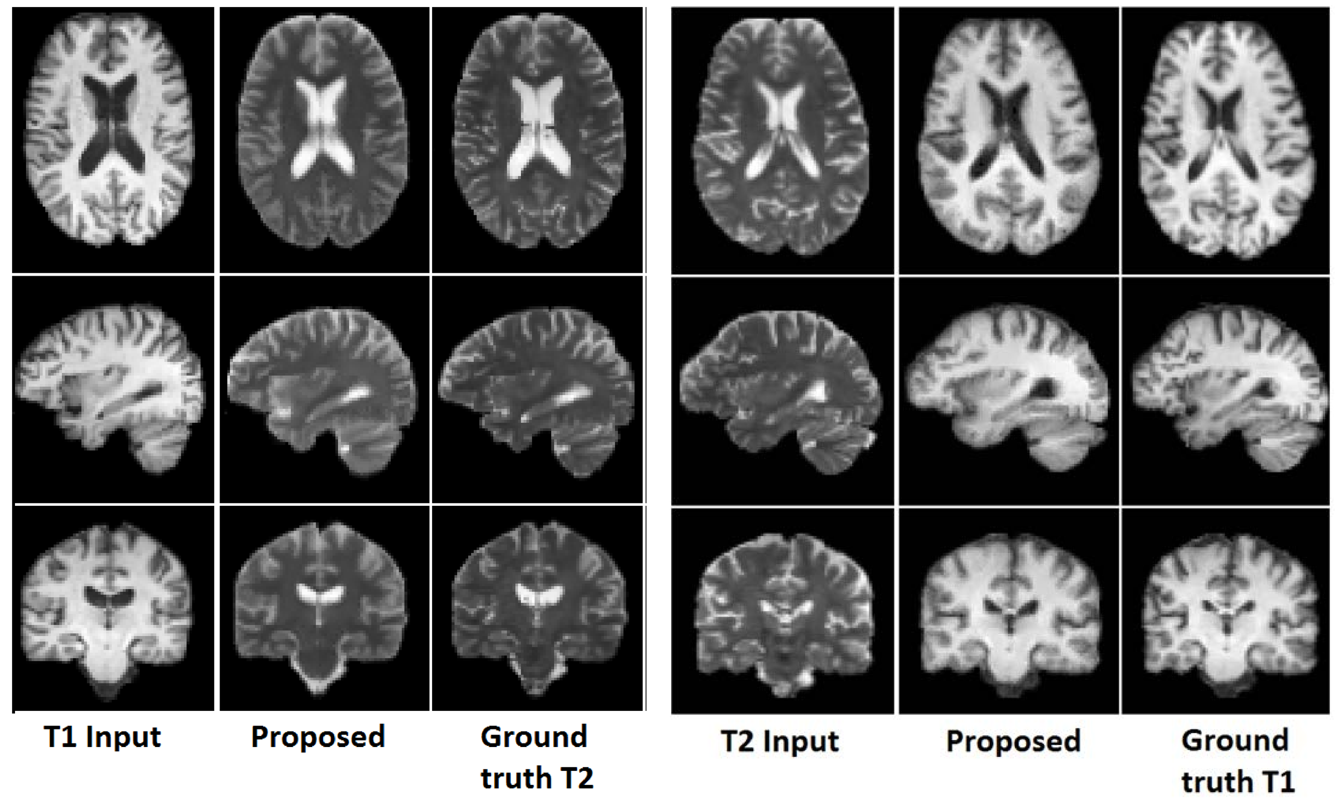
Results

- Unsupervised synthesis results:

Measure	Source /Target	F-NN	A-NN	MI Only	SC only	Proposed MI +SC	Proposed MI+SC and CSR
Correlation	T1/T2	0.717	0.815	0.808	0.809	0.839	0.855
	T2/T1	0.858	0.910	0.903	0.906	0.927	0.931
PSNR	T1/T2	10.10	12.41	11.72	12.11	12.78	13.35
	T2/T1	13.30	15.45	14.88	15.19	16.23	16.52

- Supervised synthesis results:

Measure	Source/Target	MP[2]	LSDN[3]	Proposed
Correlation	T1/T2	0.875	0.892	0.908
	T2/T1	0.931	0.941	0.953
PSNR	T1/T2	13.64	14.93	15.30
	T2/T1	15.13	17.39	18.33



References

[1] T. Cao, V. Jojic, S. Modla, D. Powell, K. Czymmek, and M. Niethammer, “Robust Multimodal Dictionary Learning”, In MICCAI, 2013.
[2] D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, and E. Konukoglu, “Modality Propagation: Coherent Synthesis of Subject-specific Scans with Data-driven Regularization”, In MICCAI, 2013.
[3] H. V. Nguyen, S. K. Zhou, and R. Vemulapalli, “Cross-Domain Synthesis of Medical Images Using Efficient Location-Sensitive Deep Network”, In MICCAI, 2015.