This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Response to the Review of TCBBSI-2019-10-0485

Dear Prof. Zhang and the Reviewers,

Title: A Novel Negative-Transfer-Resistant Fuzzy Clustering Model with a Shared Cross-Domain Transfer Latent Space and its Application to Brain CT Image Segmentation

First of all, we would like to thank the reviewers for all their positive comments. According to your comments and suggestions, we have carefully revised the manuscript. Now, we are submitting the revised manuscript (the original manuscript No. TCBBSI-2019-10-0485). Main changes are highlighted in YELLOW in the revised manuscript and the details in response to the comments are given below.

With our best efforts, we believe that the quality of the revised manuscript is okay now.

Thanks a lot in advance for reviewing the manuscript again.

Best regards

All authors

2019/12/14

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Response to Associate Editor

Comments to the Author: This paper needs a major revision according to the reviewers' comments. Some descriptions are unclear and should be clarified, such as the motivation of your proposed method and the framework of the adopted algorithm. The selection of the parameter values and their influence on the modeling performance should be discussed. In the comparison of different models, more datasets should be used to further test the performance of your method.

Reply: We would like to thank all of you for your valuable comments which have been very helpful to improve the quality of our manuscript. We have carefully revised the manuscript according to the comments received. Main changes are in yellow in the revised manuscript and the details in response to the comments are given below.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Response to Reviewer 1

This paper proposes a new image segmentation LSS-FTC-NTR model for leveraging source knowledge to improve the segmentation performance of target domain. The proposed model is interesting and the experimental results are effective. This paper looks nice. This paper can be considered for publication after a major revision. The authors should revise the paper by taking the reviewer's comments into account.

Comments 1: I think that deep learning approaches can be mentioned in Section 1 with respect to clustering performance maximization.

Reply: Thanks for your valuable comment. As suggested by the reviewer, we have revised the Introduction section in our revised manuscript, and we have added more related work in it. **(*Ref to: the first paragraph of page 2*)**:

Deep learning learns the feature representation of tissue contour based on deep convolutional neural networks [12, 13]. Deep learning methods have successfully applied for medical image segmentation in recent years. However, deep learning methods usually need a large number of training dataset and special hardware devices.

Comments 2: The number of clustering and parameter setting are important factors which are directly related to the clustering performance. In addition to the grid-type selection, some clustering index can be used. This issue can be also discussed in the paper.

Reply: Thanks for your suggestions. We agree with your opinion. In addition to the grid-type selection, some cluster validity indices can be used for determination the number of clustering and other parameters, such as Xie-Beni index, Mountain potential index and so on. In the experiments, we perform our experiments on ultrashort echo time (UTE) and modified Dixon brain image datasets. All CT images with corresponding manual segmentation are segmented into three classes: bone, water and soft issues. The number of clustering is manually set to be three, and the other parameters are determined by the grid-type selection strategy. In our future work, we will extend our work to other medical image segmentation applications. How to set the number of clustering and other parameters is worthy to be studied in the future.

Comments 3: The parameter specifications are not clearly explained. It is recommended that the author examine all the formulas throughout.

Reply: Thanks a lot for your reminder. In the revised manuscript, we have double-checked the mathematical notations and parameter specifications.

Comments 4: Some sentences are complicated and difficult to read. I personally suggest that sometimes the author doesn't need to use these long sentences.

Reply: Thanks a lot for your reminder. We have carefully proofed the manuscript to correct grammatical errors. We also have asked a technical writer to polish the manuscript. We believe the quality of the revised manuscript has been improved significantly.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Comments 5: It would be interesting if the proposed model is compared on more medical datasets.

Reply: Thanks for your suggestions. In response to both your comment and another referee's suggestions, we do changes in the revised manuscript as follows

1) We have performed more experiments in our revised manuscript. **(*Ref to: the subsection 4.1 of page 7*)**:

We randomly select 20 brain CT images as the original target domain data, and the rest 236 brain CT images as source domain data. Following the training protocol established in [41], we construct a total training data set combining 236 source brain images and random 8 target brain images, while the remaining 12 target brain images are used as testing brain images. We repeat the experiment for 10 runs and record the experimental results.

Table 1

NMI performance of all comparison methods on 5% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0112	0.0041	0.5085	0.2249	0.5802	0.0254	**0.7146**	0.0056
Subject2	0.1515	0.1468	0.5801	0.1632	0.6359	0.0148	**0.7526**	0.0142
Subject3	0.1583	0.1040	0.5573	0.1591	0.6438	0.0290	**0.7253**	0.0059
Subject4	0.0137	0.0062	0.5149	0.1868	0.6091	0.0590	**0.7256**	0.0138
Subject5	0.2333	0.1355	0.5649	0.1840	0.6329	0.0254	**0.7432**	0.0094
Subject6	0.1823	0.1094	0.5659	0.1446	0.6505	0.0234	**0.7441**	0.0116
Subject7	0.0086	0.0046	0.5283	0.2242	0.6096	0.0212	**0.7352**	0.0110
Subject8	0.1324	0.1157	0.5664	0.1917	0.6603	0.0245	**0.7586**	0.0108
Subject9	0.1188	0.1240	0.5845	0.1638	0.6700	0.0399	**0.7693**	0.0111
Subject10	0.0185	0.0083	0.5765	0.2749	0.6512	0.0258	**0.7781**	0.0127
Subject11	0.1178	0.2362	0.5882	0.2124	0.6580	0.0179	**0.7862**	0.0068
Subject12	0.0642	0.0376	0.5719	0.1542	0.6631	0.0358	**0.7759**	0.0079

Table 2

NMI performance of all comparison methods on 10% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0155	0.0098	0.5063	0.2137	0.4748	0.0095	**0.6337**	0.0091
Subject2	0.2943	0.1666	0.5797	0.1689	0.5792	0.0194	**0.7230**	0.0045
Subject3	0.2182	0.0813	0.5502	0.1716	0.5224	0.0366	**0.6908**	0.0117
Subject4	0.0196	0.0163	0.5061	0.1966	0.4706	0.0110	**0.6475**	0.0127
Subject5	0.1959	0.1420	0.5635	0.1591	0.5273	0.0205	**0.7056**	0.0122
Subject6	0.2868	0.0777	0.5683	0.1421	0.5310	0.0162	**0.6894**	0.0177
Subject7	0.0076	0.0038	0.5336	0.2612	0.4696	0.0106	**0.6710**	0.0162
Subject8	0.2006	0.1377	0.5680	0.1448	0.5357	0.0217	**0.7077**	0.0158
Subject9	0.2052	0.1502	0.5923	0.1901	0.5409	0.0165	**0.7274**	0.0104
Subject10	0.0153	0.0136	0.5755	0.2609	0.5103	0.0186	**0.7244**	0.0046
Subject11	0.0814	0.1664	0.5846	0.2263	0.5231	0.0215	**0.7203**	0.0122
Subject12	0.1534	0.1036	0.5694	0.1864	0.5409	0.0062	**0.7144**	0.0149

Table 3

NMI performance of all comparison methods on 15% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0130	0.0106	0.5053	0.2261	0.4470	0.0181	**0.5684**	0.0082
Subject2	0.2346	0.1282	0.5815	0.1536	0.5631	0.0172	**0.6452**	0.0118
Subject3	0.2155	0.0679	0.5541	0.2003	0.4881	0.0175	**0.5978**	0.0024
Subject4	0.1789	0.2392	0.5076	0.2182	0.4509	0.0171	**0.5849**	0.0157
Subject5	0.2061	0.0862	0.5621	0.1394	0.5341	0.0121	**0.6220**	0.0030
Subject6	0.2592	0.1368	0.5691	0.1526	0.5079	0.0163	**0.6261**	0.0096
Subject7	0.0963	0.1848	0.5274	0.1991	0.4664	0.0168	**0.5939**	0.0130
Subject8	0.1826	0.1223	0.5674	0.1458	0.5087	0.0223	**0.6198**	0.0101
Subject9	0.2245	0.1570	0.5869	0.1486	0.5118	0.0120	**0.6419**	0.0157
Subject10	0.0735	0.1450	0.5697	0.2271	0.4975	0.0197	**0.6098**	0.0169
Subject11	0.0937	0.1846	0.5826	0.2770	0.5267	0.0242	**0.6107**	0.0075
Subject12	0.2454	0.1816	0.5720	0.2135	0.5328	0.0129	**0.6537**	0.0117

Table 4

NMI performance of all comparison methods on 20% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0246	0.0289	0.5095	0.1908	0.4782	0.0069	**0.6000**	0.0055
Subject2	0.3092	0.1299	0.5796	0.1631	0.5596	0.0082	**0.6348**	0.0128
Subject3	0.2597	0.0992	0.5556	0.1588	0.5029	0.0108	**0.5926**	0.0051
Subject4	0.0102	0.0061	0.5136	0.2026	0.4778	0.0066	**0.5254**	0.0133
Subject5	0.3071	0.0852	0.5654	0.1471	0.5333	0.0119	**0.6237**	0.0084
Subject6	0.2998	0.0697	0.5662	0.1376	0.5247	0.0176	**0.5875**	0.0099
Subject7	0.0994	0.1742	0.5298	0.2377	0.4804	0.0088	**0.5719**	0.0063
Subject8	0.2557	0.1427	0.5683	0.1438	0.5426	0.0160	**0.6023**	0.0114
Subject9	0.3098	0.0342	0.5852	0.1508	0.5366	0.0089	**0.6433**	0.0062
Subject10	0.0966	0.2007	0.5788	0.2565	0.5309	0.0066	**0.5991**	0.0080
Subject11	0.1196	0.2075	0.5818	0.2101	0.5525	0.0080	**0.6516**	0.0129
Subject12	0.1530	0.1197	0.5707	0.1767	0.5466	0.0167	**0.5869**	0.0122

Table 5

NMI performance of all comparison methods on 25% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0107	0.0058	0.5078	0.2154	0.4969	0.0043	**0.5469**	0.0104
Subject2	0.1708	0.1001	0.5792	0.1741	0.5842	0.0076	**0.6572**	0.0182
Subject3	0.2460	0.0981	0.5503	0.1685	0.5232	0.0145	**0.5940**	0.0093
Subject4	0.0121	0.0081	0.5177	0.2390	0.5045	0.0112	**0.5717**	0.0197

Subject5	0.1503	0.1169	0.5621	0.1599	0.5580	0.0122	**0.6090**	0.0083
Subject6	0.2199	0.0660	0.5662	0.1533	0.5434	0.0311	**0.5935**	0.0142
Subject7	0.0284	0.0285	0.5331	0.2528	0.5156	0.0137	**0.5677**	0.0095
Subject8	0.1320	0.1375	0.5705	0.1577	0.5547	0.0156	**0.6272**	0.0136
Subject9	0.2594	0.1054	0.5859	0.1542	0.5682	0.0124	**0.6437**	0.0127
Subject10	0.1980	0.0293	0.5763	0.2203	0.5488	0.0073	**0.6416**	0.0157
Subject11	0.2156	0.2697	0.5830	0.2349	0.5745	0.0184	**0.6732**	0.0117
Subject12	0.3935	0.0908	0.5724	0.1827	0.5617	0.0122	**0.6386**	0.0078

Table 6

NMI performance of all comparison methods on 30% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0120	0.0093	0.5086	0.2028	0.5569	0.0124	**0.6567**	0.0084
Subject2	0.2553	0.1770	0.5833	0.1497	0.6333	0.0113	**0.7439**	0.0109
Subject3	0.2010	0.0804	0.5526	0.1537	0.5872	0.0206	**0.6602**	0.0110
Subject4	0.0891	0.1752	0.5143	0.2144	0.5664	0.0050	**0.6720**	0.0114
Subject5	0.1820	0.1376	0.5663	0.1463	0.6172	0.0193	**0.7080**	0.0028
Subject6	0.2316	0.1094	0.5695	0.1928	0.6062	0.0120	**0.6872**	0.0086
Subject7	0.1080	0.1686	0.5316	0.2091	0.5769	0.0172	**0.6966**	0.0054
Subject8	0.1755	0.1045	0.5696	0.1777	0.6231	0.0097	**0.7037**	0.0103
Subject9	0.2584	0.0893	0.5913	0.1901	0.6295	0.0216	**0.7356**	0.0200
Subject10	0.0113	0.0135	0.5751	0.2901	0.6247	0.0198	**0.7441**	0.0115
Subject11	0.2121	0.2838	0.5827	0.2652	0.6343	0.0212	**0.7491**	0.0042
Subject12	0.1547	0.1170	0.5702	0.1941	0.6226	0.0045	**0.7256**	0.0095

Table 7

ARI performance of all comparison methods on 5% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0021	0.0043	0.3399	0.1756	0.7579	0.0301	**0.8880**	0.0028
Subject2	0.0854	0.0814	0.3931	0.1561	0.8032	0.0160	**0.9111**	0.0109
Subject3	0.0861	0.0521	0.4144	0.1478	0.8072	0.0318	**0.8839**	0.0031
Subject4	0.0013	0.0041	0.3466	0.1701	0.7730	0.0629	**0.8913**	0.0107
Subject5	0.1186	0.0877	0.3884	0.1761	0.8085	0.0316	**0.9080**	0.0078
Subject6	0.0839	0.0560	0.4153	0.1393	0.8069	0.0264	**0.8950**	0.0108
Subject7	0.0012	0.0026	0.3593	0.1594	0.7728	0.0243	**0.8922**	0.0119
Subject8	0.0651	0.0732	0.3902	0.1753	0.8293	0.0276	**0.9169**	0.0095
Subject9	0.0658	0.0682	0.4153	0.1524	0.8305	0.0378	**0.9187**	0.0102
Subject10	0.0071	0.0055	0.3806	0.2090	0.8036	0.0304	**0.9242**	0.0106
Subject11	0.0771	0.1614	0.3858	0.1903	0.8089	0.0215	**0.9299**	0.0053
Subject12	0.0094	0.0109	0.3744	0.1365	0.8131	0.0424	**0.9269**	0.0081

Table 8

ARI performance of all comparison methods on 10% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0028	0.0113	0.3403	0.1791	0.5742	0.0284	**0.8155**	0.0056
Subject2	0.1798	0.1182	0.3955	0.1613	0.5515	0.0292	**0.8812**	0.0056
Subject3	0.1158	0.0542	0.4121	0.1696	0.6163	0.0635	**0.8515**	0.0148
Subject4	0.0046	0.0063	0.3433	0.1729	0.5650	0.0237	**0.8281**	0.0134
Subject5	0.1051	0.0953	0.3848	0.1535	0.6339	0.0317	**0.8777**	0.0093
Subject6	0.1523	0.0812	0.4168	0.1373	0.6359	0.0297	**0.8467**	0.0136
Subject7	0.0008	0.0012	0.3607	0.1920	0.5573	0.0211	**0.8482**	0.0125
Subject8	0.1140	0.0988	0.3915	0.1311	0.6423	0.0301	**0.8522**	0.0149
Subject9	0.1202	0.1143	0.4210	0.1781	0.6382	0.0300	**0.8839**	0.0101
Subject10	0.0011	0.0089	0.3800	0.1963	0.5939	0.0328	**0.8849**	0.0037
Subject11	0.0493	0.1062	0.3848	0.1608	0.6045	0.0401	**0.8742**	0.0118
Subject12	0.0484	0.0401	0.3737	0.1681	0.6302	0.0112	**0.8738**	0.0125

Table 9

ARI performance of all comparison methods on 15% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0009	0.0031	0.3389	0.1753	0.3607	0.0262	**0.6833**	0.0089
Subject2	0.1206	0.1058	0.3935	0.1414	0.4404	0.0126	**0.6679**	0.0107
Subject3	0.1083	0.0540	0.4127	0.1807	0.4323	0.0238	**0.5997**	0.0032
Subject4	0.1239	0.1680	0.3439	0.1746	0.3620	0.0344	**0.7034**	0.0137
Subject5	0.0847	0.0405	0.3859	0.1306	0.4408	0.0298	**0.7278**	0.0046
Subject6	0.1421	0.1255	0.4182	0.1430	0.4259	0.0204	**0.6636**	0.0077
Subject7	0.0639	0.1304	0.3588	0.1519	0.3662	0.0287	**0.5863**	0.0161
Subject8	0.0729	0.0824	0.3902	0.1349	0.4120	0.0355	**0.5671**	0.0136
Subject9	0.1302	0.0900	0.4174	0.1444	0.4186	0.0102	**0.6287**	0.0148
Subject10	0.0398	0.0848	0.3778	0.2008	0.4098	0.0241	**0.5256**	0.0122
Subject11	0.0593	0.1239	0.3843	0.2049	0.4230	0.0250	**0.6560**	0.0087
Subject12	0.1478	0.1185	0.3750	0.1692	0.4119	0.0303	**0.6518**	0.0128

Table 10

ARI performance of all comparison methods on 20% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0019	0.0080	0.3400	0.1486	0.3241	0.0059	**0.6624**	0.0098
Subject2	0.1754	0.1208	0.3933	0.1522	0.3859	0.0059	**0.6005**	0.0116
Subject3	0.1443	0.0831	0.4147	0.1533	0.3742	0.0102	**0.5570**	0.0066
Subject4	0.0015	0.0060	0.3464	0.1713	0.3270	0.0090	**0.5008**	0.0139
Subject5	0.1679	0.0931	0.3883	0.1370	0.3730	0.0056	**0.6152**	0.0068

Subject6	0.1701	0.0682	0.4176	0.1304	0.3838	0.0091	**0.5775**	0.0079
Subject7	0.0625	0.1226	0.3602	0.1912	0.3309	0.0064	**0.4983**	0.0083
Subject8	0.1520	0.1066	0.3913	0.1339	0.3809	0.0088	**0.5055**	0.0102
Subject9	0.1619	0.0404	0.4164	0.1424	0.3871	0.0056	**0.5725**	0.0053
Subject10	0.0620	0.1394	0.3812	0.1927	0.3571	0.0027	**0.5494**	0.0065
Subject11	0.0713	0.1457	0.3838	0.1649	0.3695	0.0065	**0.5819**	0.0122
Subject12	0.0657	0.0780	0.3746	0.1617	0.3641	0.0048	**0.4376**	0.0106

Table 11

ARI performance of all comparison methods on 25% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0023	0.0020	0.3409	0.1786	0.3419	0.0056	**0.5128**	0.0139
Subject2	0.0687	0.0687	0.3948	0.1626	0.4354	0.0132	**0.6609**	0.0156
Subject3	0.1447	0.0664	0.4126	0.1659	0.4043	0.0113	**0.5219**	0.0102
Subject4	0.0125	0.0047	0.3484	0.1856	0.3738	0.0138	**0.5349**	0.0204
Subject5	0.0701	0.0553	0.3862	0.1545	0.4182	0.0235	**0.6445**	0.0077
Subject6	0.1147	0.0392	0.4186	0.1502	0.4153	0.0305	**0.5271**	0.0151
Subject7	0.0076	0.0109	0.3611	0.1885	0.3654	0.0088	**0.5132**	0.0116
Subject8	0.0656	0.0666	0.3914	0.1499	0.4143	0.0176	**0.5902**	0.0148
Subject9	0.1248	0.0841	0.4173	0.1462	0.4233	0.0152	**0.6459**	0.0121
Subject10	0.1157	0.0035	0.3804	0.1958	0.3796	0.0046	**0.5953**	0.0165
Subject11	0.1382	0.1889	0.3842	0.1686	0.4277	0.0200	**0.6537**	0.0123
Subject12	0.2421	0.0869	0.3739	0.1749	0.4058	0.0177	**0.6031**	0.0091

Table 12

ARI performance of all comparison methods on 30% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0008	0.0052	0.3401	0.1505	0.5432	0.0294	**0.7743**	0.0062
Subject2	0.1663	0.1180	0.3941	0.1505	0.5887	0.0309	**0.8122**	0.0136
Subject3	0.0947	0.0389	0.4132	0.1584	0.5700	0.0397	**0.7414**	0.0125
Subject4	0.0601	0.1242	0.3471	0.1760	0.5526	0.0094	**0.7876**	0.0146
Subject5	0.0867	0.0967	0.3886	0.1351	0.5858	0.0488	**0.7966**	0.0044
Subject6	0.1193	0.0774	0.4173	0.1744	0.5833	0.0131	**0.7834**	0.0053
Subject7	0.0675	0.1173	0.3611	0.1547	0.5637	0.0408	**0.7874**	0.0079
Subject8	0.0684	0.0490	0.3911	0.1728	0.6018	0.0277	**0.7363**	0.0144
Subject9	0.1227	0.0787	0.4184	0.1823	0.6060	0.0214	**0.8233**	0.0176
Subject10	0.0047	0.0076	0.3800	0.2041	0.5840	0.0421	**0.8140**	0.0134
Subject11	0.1420	0.1947	0.3839	0.2030	0.6010	0.0318	**0.8226**	0.0052
Subject12	0.0588	0.0640	0.3745	0.1792	0.5793	0.0106	**0.7996**	0.0093

2) We perform the application of LSS-FTC-NTR in the scenario of target images polluted by 20%,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

25% and 30% Gaussian noise in our revised manuscript. To better observe the behavior of all algorithms, Figs. 4-9 graphically shows the segmentation results of all comparison methods obtained on subject1 with different noise. Similar to the results in the Tables 1-12, LSS-FTC-NTR obtains the best segmentation results for distinguishing the bone, water and soft issues. The boundaries between different organizations are smooth and obvious are relatively clearer than the other three methods.

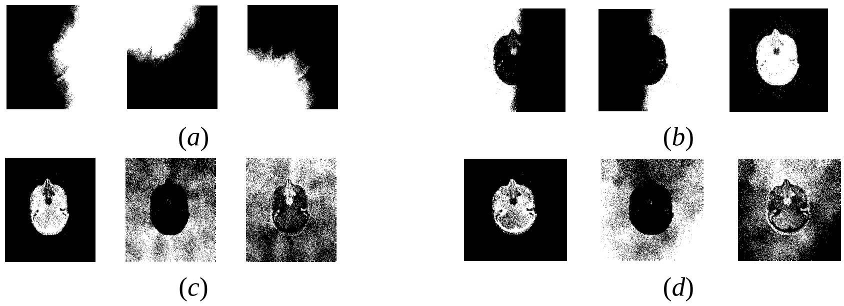

Fig.7 Clustering segmentations on subject1+20% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

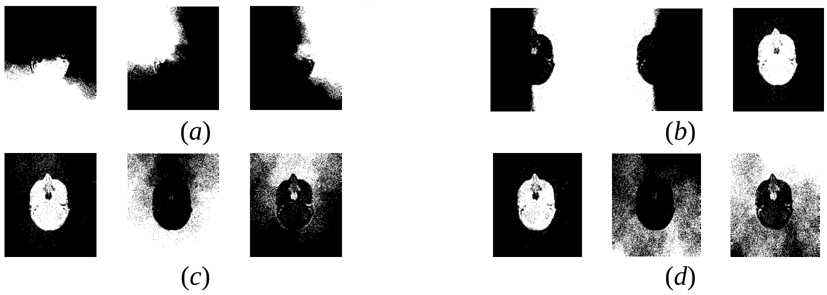

Fig.8 Clustering segmentations on subject1+25% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

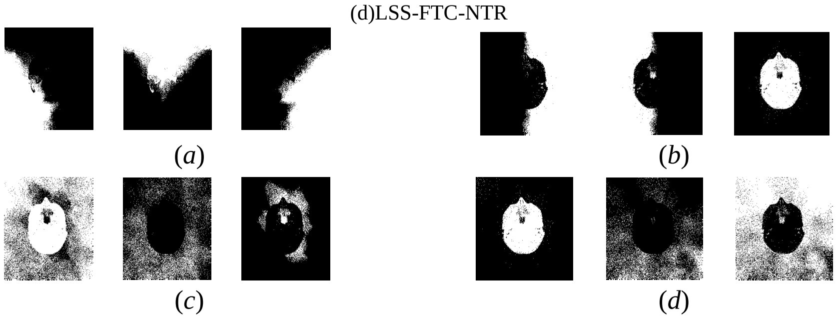

Fig.9 Clustering segmentations on subject1+30% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

Comments 6: The discrepancies of the results in the experiments are easy to understand. However, there lacks necessary explanation about how the proposed model outperforms others. This makes it inadequate to assert the effectiveness of the proposed model. Please give a more detailed explanation of the results.

Reply: We would like to thank you for your comments. In order to follow your suggestions, firstly we have restated the information of datasets used in the experiment (***Please refer to the revision in the Section 4.1, page 7***), then we have improved the presentation of the experimental results in revised manuscript (***Please refer to the revision in the Section 4.2, pages 8-10***). Moreover, we have given a more detailed explanation about the experimental results (***Please refer to the revision in the Section 4.2, page 8***).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Response to Reviewer 2

Comments 1: The study proposes a negative-transfer-resistant mechanism by using the weight of transferred knowledge to achieve positive transfer and avoid negative transfer. In addition, the integrated negative-transfer-resistant and maximum mean discrepancy into the framework of fuzzy c-means clustering is also proposed in this manuscript. However, I personally believe that this study is not yet accepted. Some minor revisions are required.

Reply: Thank you for your comments. We have carefully revised the manuscript according to the comments received. Main changes are in yellow in the revised manuscript and the details in response to the comments are given below.

Comments 2: The authors consider the noisy scenario as the target domain and the existing medical image dataset from related scenario elsewhere as the source domain, and use the learning on clean images of source data to improve the clustering in target data, but I personally think that the contributions and research motivations of this study should be clearly described in the abstract.

Reply: Thank you for your comments. In order to follow your suggestions, we have made the following changes in the revised manuscript:

1) We have added up the motivation of LSS-FTC-NTR in Section 1 (***Ref to: the first paragraph of page 3***):

The motivation of LSS-FTC-NTR is shown in Fig.1. Two cluster centers presented as black triangle and circle have positive transfer influence to the clustering in the target domain, while the cluster center presented as Black Square has negative influence to the clustering in the target domain. LSS-FTC-NTR will automatically resist black square participating in the clustering in the target domain by using the negative-transfer-resistance strategy.

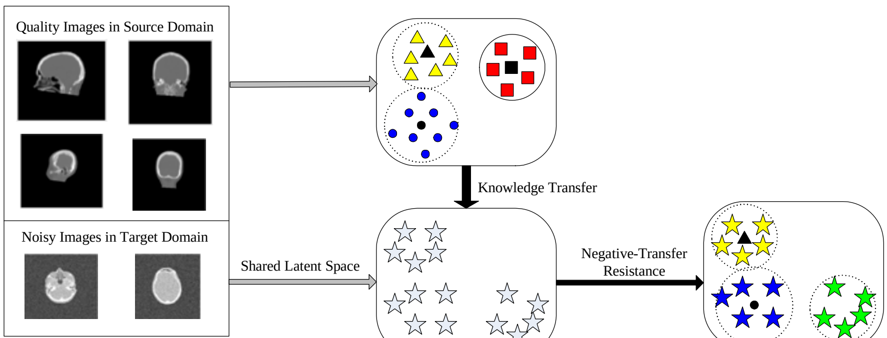

Fig.1 The motivation of LSS-FTC-NTR

2) We have added up the contributions of LSS-FTC-NTR in Section 1 (***Ref to: the second paragraph of page 3***):

The novelty of this study is as follows.

1) We formulate the problem of insufficient and noisy medical image segmentation as a model of transfer clustering task. To the best of our knowledge, our study is the first attempt to address this issue.

2) The negative-transfer-resistance mechanism is proposed to identify and resist negative source transfer knowledge.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

3) The MMD is introduced into LSS-FTC-NTR to unify the representation of image data of different domains in the shared latent space, which helps transferring knowledge across domains.
4) Clustering centers based transfer matching scheme is used to deal with the inconsistency problem of clustering numbers between source and target domains, so that more robust and cluster performance can be promoted.

Comments 3: Also, the author should clearly point out the differences and relationships between the research and the existing related work.
Reply: Thank you for your valuable comments. We have added some discussion to compare transfer learning with multi-task learning and co-clustering. **(*Ref to: the last paragraph of page 3*)**:
Currently, when the training data is not enough to represent the current domain, transfer learning, multi-task learning and co-clustering are three effective techniques that can enhance the clustering performance in the current domain. Multi-task learning performs multiple learning tasks together through by sharing certain knowledge among all tasks [30, 31]. Co-clustering performs clustering on both rows and columns of data samples, so as to exploit the clear duality between rows and columns of a contingency table [32]. Transfer learning clustering enhances the clustering performance in the new domain by leveraging useful knowledge from different but related domains. Many researches show that the transfer clustering methods have better learning ability to obtain an effective model with the idea of transfer learning [28, 29, 33]. In real applications, due to the existences of noise and field offset etc, the insufficient medical images are inadequate to complete image segmentation. Therefore, we think transfer learning clustering is an effect technology to promote the segmentation of insufficient and noisy medical image in the new domain.

Comments 4: It is very important that, in order to better use the method for readers, the personal recommendation should indicate the actual application scenario and the issues to be considered in the conclusion.
Reply: Thank you for your comments. In actual application scenario, medical images are often collected with different scanners and scanning parameters, medical images may have large differences in image quality due to machine performance or scanning technology, such as varying degrees of rotation, noise, etc. The requirement of training and target data under the same distribution prevents the use of clustering algorithms in larger research and clinical practice. Since the above scenarios exist in a large number of real-world environments, this leads to unsatisfactory segmentation results and the risk of algorithm failure. In this paper, we study the problem of medical image segmentation in a noisy scenario by transferring medical images collected from related scenarios. We consider the new noisy scenario as the target domain and the existing medical image dataset from related scenario elsewhere as the source domain, and then use the learning on clean images of source data to improve the clustering in target data. To improve the transfer learning performance, we consider learning the negative-transfer-resistant mechanism, so that the influence of positive transfer knowledge is reinforced and the influence of negative transfer knowledge is reduced or even eliminated. Meanwhile, we think medical images in different scenarios share certain common representations such as bone and soft tissue,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

and the shared representations could be preserved in a shared space. In the revised manuscript, we indicate this issue and the future work in the conclusion. (***Ref to: the conclusion subsection of page 13***):

The experiments focus on noisy brain CT images. The experimental results show that with insufficient and noisy medical images, it is possible to build an efficient segmentation model with the help of medical images from the related scenarios. Future work will extend our algorithm to other medical image segmentation applications. We will extend the framework so as to apply various clustering algorithms in order to obtain more satisfactory medical image segmentation results. We will also study how many images in the source domain can be considered sufficient, and how to select the important images to further improve the transfer. In addition, how to speed up LSS-FTC-NTR is worthy to be studied in the future.

Comments 5: Although the study proposes an optimization of LSS-FTC-NTR In sub-section 3.3, the author should briefly analyze and discuss the proposed algorithm.

Reply: Thank you for your comments. In order to follow your suggestions, we have made the following changes in the revised manuscript:

1) We have added up the motivation of LSS-FTC-NTR in Section 1 (***Ref to: the first paragraph of page 3***):

The motivation of LSS-FTC-NTR is shown in Fig.1. Two cluster centers presented as black triangle and circle have positive transfer influence to the clustering in the target domain, while the cluster center presented as Black Square has negative influence to the clustering in the target domain. LSS-FTC-NTR will automatically resist black square participating in the clustering in the target domain by using the negative-transfer-resistance strategy.

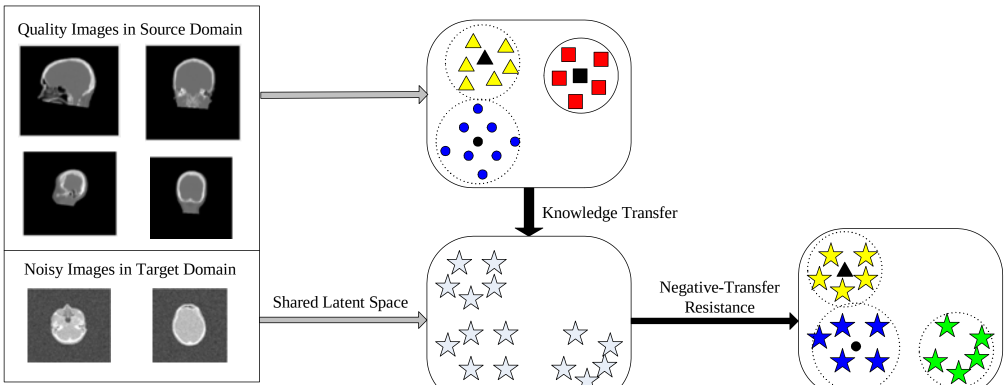

Fig.1 The motivation of LSS-FTC-NTR

2) We have added up the novelty of LSS-FTC-NTR in Section 1 (***Ref to: the second paragraph of page 3***):

The novelty of this study is as follows. 1) We formulate the problem of insufficient and noisy medical image segmentation as a model of transfer clustering task. To the best of our knowledge, our study is the first attempt to address this issue. 2) The negative-transfer-resistance mechanism is proposed to identify and resist negative source transfer knowledge. 3) The MMD is introduced into LSS-FTC-NTR to unify the representation of image data of different domains in the shared latent space, which helps transferring knowledge across domains. 4) Clustering centers based transfer matching scheme is used

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

to deal with the inconsistency problem of clustering numbers between source and target domains, so that more robust and cluster performance can be promoted.

3) To clearly describe the algorithm of LSS-FTC-NTR, we have re-written some steps of Algorithm1 (***Ref to: the first paragraph of page 6***):

Algorithm 1: LSS-FTC-NTR model	
Initialize	Set the maximum number of iterations t_{max}, the fuzzy index m, the regularization parameters λ_1 and λ_2, and the learning rate η .
Repeat:	
Exacting transfer knowledge form the source domain;	
Perform soft-partition clustering methods in the source domain, such as FCM, and obtain the cluster centers of data in the source domain;	
$t = t+1$;	
Initialize the clustering centers of data in the target domain;	
Compute the weight of transfer knowledge S_{jh} using Eq. (10);	
Fix $\mathbf{U}(t)$ and $\mathbf{\Theta}(t)$, obtain $\tilde{\mathbf{V}}^{TD}(t)$ using Eq. (12);	
Fix $\tilde{\mathbf{V}}^{TD}(t)$ and $\mathbf{\Theta}(t)$, obtain $\mathbf{U}(t)$ using Eq. (14);	
Fix $\mathbf{U}(t)$ and $\tilde{\mathbf{V}}^{TD}(t)$, obtain $\mathbf{\Theta}(t)$ using Eq. (18) and Eq.(19);	
Compute $J(t)$ using Eq. (9);	
Until $\|J(t)-J(t-1)\| \le \delta$ or $t \ge t_{max}$;	

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Response to Reviewer 3

Comments 1: In this paper, authors propose a negative-transfer-resistance fuzzy clustering model with a shared cross-domain transfer latent space. In this paper, ultrashort echo time (UTE) and modified Dixon brain image datasets are considered. However, the below few points are unclear to the referees:

The specific arrangements of the manuscript is structured as follows: The first part is an introduction, describes the purpose and significance of this study of the problem and the background, literature review, research methods and thesis structure arrangements. Organization of paper is okay. I think you should add more previous research in the field of medical image segmentation.

Reply: Thank you for your comments. In order to follow your suggestions, we have made the following changes in the revised manuscript:

1) We have added some discussion about deep learning in the Introduction section. **(*Ref to: the second paragraph of page 3*)**:

Deep learning learns the feature representation of tissue contour based on deep convolutional neural networks [12, 13]. Deep learning methods have successfully applied for medical image segmentation in recent years. However, deep learning methods usually need a large number of training dataset and special hardware devices.

2) We have added some discussion to compare transfer learning with multi-task learning and co-clustering. **(Ref to: the second paragraph of page 3)**:

Currently, when the training data is not enough to represent the current domain, transfer learning, multi-task learning and co-clustering are three effective techniques that can enhance the clustering performance in the current domain. Multi-task learning performs multiple learning tasks together through by sharing certain knowledge among all tasks [30, 31]. Co-clustering performs clustering on both rows and columns of data samples, so as to exploit the clear duality between rows and columns of a contingency table [32]. Transfer learning clustering enhances the clustering performance in the new domain by leveraging useful knowledge from different but related domains. Many researches show that the transfer clustering methods have better learning ability to obtain an effective model with the idea of transfer learning [28, 29, 33]. In real applications, due to the existences of noise and field offset etc, the insufficient medical images are inadequate to complete image segmentation. Therefore, we think transfer learning clustering is an effect technology to promote the segmentation of insufficient and noisy medical image in the new domain.

2) We have added some relevant introduction work about FCM-based transfer learning methods **(*Ref to: the last second paragraph of page 3*)**:

For example, a FCM-based transfer learning was proposed in [28], which is combined with Gini-Simpson diversity index and quadratic weights on membership. A knowledge-leveraged transfer FCM (KL-TFCM) is proposed in [29], which uses three-interlinked framework of knowledge extraction, knowledge matching, and knowledge utilization to leverage source information to help the FCM clustering in the target domain. However, the above FCM-based transfer learning clustering methods are completed in the original space, and while they do not consider resisting negative transfer.

Comments 2: I suggest the authors give the motivation of the paper in the first part, which is convenient for readers to understand the whole framework of the paper.

Reply: We would like to thank you for your comments. In order to follow your

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

suggestions, we have added up the motivation of LSS-FTC-NTR in Section 1 (***Ref to: the first paragraph of page 3***):

The motivation of LSS-FTC-NTR is shown in Fig.1. Two cluster centers presented as black triangle and circle have positive transfer influence to the clustering in the target domain, while the cluster center presented as Black Square has negative influence to the clustering in the target domain. LSS-FTC-NTR will automatically resist black square participating in the clustering in the target domain by using the negative-transfer-resistance strategy.

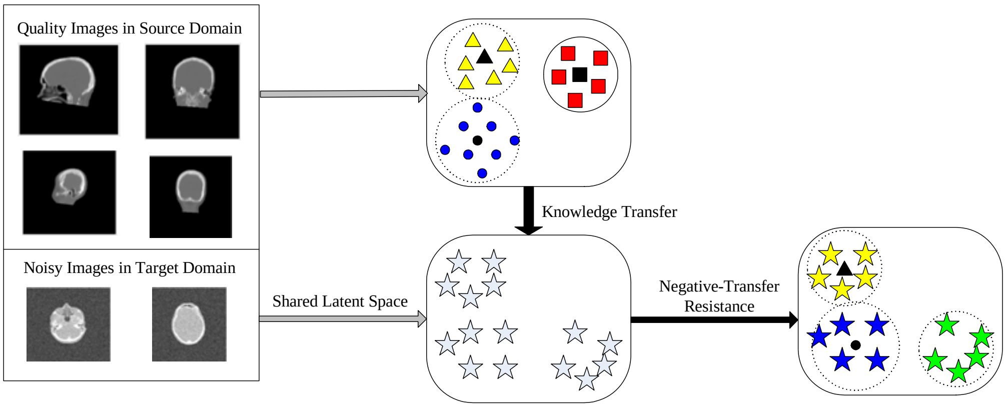

Fig.1 The motivation of LSS-FTC-NTR

Comments 3: Figure 1 is a good description of the proposed schematic diagram, but the analysis of this figure is still not sufficient. The authors should pay more attention to the comparisons for the superiority or novelty of your method.

Reply: Thank you for your comments. As suggested by the reviewer, we have added the discussion about the novelty of our study in section 1. (R***ef to: the second paragraph of page 3***):

The novelty of this study is as follows.

1) We formulate the problem of insufficient and noisy medical image segmentation as a model of transfer clustering task. To the best of our knowledge, our study is the first attempt to address this issue.

2) The negative-transfer-resistance mechanism is proposed to identify and resist negative source transfer knowledge.

3) The MMD is introduced into LSS-FTC-NTR to unify the representation of image data of different domains in the shared latent space, which helps transferring knowledge across domains.

4) Clustering centers based transfer matching scheme is used to deal with the inconsistency problem of clustering numbers between source and target domains, so that more robust and cluster performance can be promoted.

Comments 4: On the basis of the research in this paper, the conclusion section details the next steps and prospects.

Reply: Thanks a lot for your positive comment.

Comments 5: Please note the format of the reference.

Reply: Thanks for your valuable comment. In our revised paper, we format the references.

Comments 6: Read your paper carefully and modify your spelling mistake throughout your manuscript.

Reply: Thanks a lot for your reminder. We have carefully proofed the manuscript to correct grammatical errors. We also have asked a technical writer to polish the manuscript. We believe the quality of the revised manuscript has been improved significantly.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Response to Reviewer 4

Comments 1: The method described in the manuscript appears to be sound and the structure is overall good. However, it needs major revision before it is accepted. My comments are as follows:

The motivation of the proposed method should be strengthened. I suggest the authors to add a figure which can be used to clearly describe your proposed method.

Reply: We would like to thank you for your comment. In response to both your comment and another referee's suggestions, in our revised manuscript, we have added up the motivation of LSS-FTC-NTR in Section 1 **(*Ref to: the second paragraph of page 3*)**:

The motivation of LSS-FTC-NTR is shown in Fig.1. Two cluster centers presented as black triangle and circle have positive transfer influence to the clustering in the target domain, while the cluster center presented as Black Square has negative influence to the clustering in the target domain. LSS-FTC-NTR will automatically resist black square participating in the clustering in the target domain by using the negative-transfer-resistance strategy.

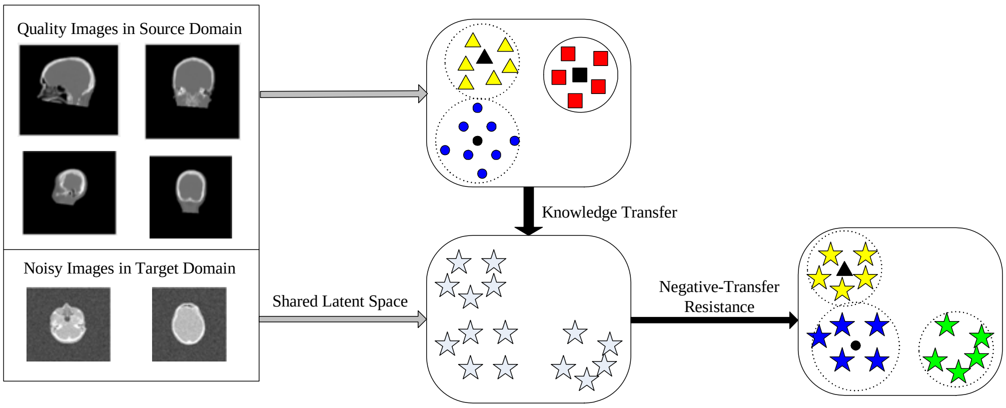

Fig.1 The motivation of LSS-FTC-NTR

Comments 2: It seems to me that the performance of your proposed method depends on these predefined parameters. Thus, I suggest you to provide a sensitivity analysis of predefined parameters.

Reply: Thank you. In the experiment, we select λ_1 and λ_2 in a given search grid. In our revised manuscript, we discuss the performance of LSS-FTC-NTR using different parameters. **(*Ref to: the subsection 4.3*)**:

Tables 15-16 show the means of NMI and ARI on the subject using different λ_1 and λ_2, while fixing the parameter m=2.

1) LSS-FTC-NTR is sensitive to parameters λ_1 and λ_2. Different λ_1 and λ_2 lend to different cluster performance of LSS-FTC-NTR in terms of NMI and ARI. It can be found that in most situations when the value of NMI is better, the value of ARI is also better. Thus, it is feasible to use NMI and ARI as performance criterions to determine the suitable parameters.

2) Fixed the value of m, LSS-FTC-NTR obtains the worst NMI and ARI when $\lambda_1 = 0$ and $\lambda_2 = 0$. The clustering performance of LSS-FTC-NTR is improved when λ_1 and λ_2 are not equal to 0. Since when $\lambda_1 = 0$ and $\lambda_2 = 0$ LSS-FTC-NTR is degenerated to the classical FCM clustering.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

3) We can find that when the value of λ_1 *is large, LSS-FTC-NTR obtains the satisfactory performance in terms of NMI and ARI. This further demonstrates that the proposed negative-transfer-resistance mechanism has played an effective role. Thus, in the subsequent experiments, we can reduce the search grid of* λ_1 *in the range* $\{10e1, 10e1, ..., 10e6\}$*. We can't find the rule to select parameter* λ_2*. We think it is reasonable to select optimal* λ_2 *within the search grid. The range* $\lambda_2 \in \{10e\text{-}4, 10e\text{-}3, ..., 10e6\}$ *is appropriate.*

Table 15 Means of NMI by LSS-FTC-NTR on the subject1+5% noise using different λ_1 and λ_2, while fixing *m*=2

λ_2 \ λ_1	0	10e-4	10e-3	10e-2	10e-1	1	10e1	10e2	10e3	10e4	10e5	10e6
0	0.4801	0.5006	0.5181	0.5501	0.6011	0.6250	0.6091	0.6375	0.6788	0.6397	0.6427	0.6378
10e-4	0.4915	0.5326	0.5592	0.5662	0.6109	0.6161	0.6469	0.6499	0.6468	0.6493	0.6431	0.6425
10e-3	0.5094	0.5436	0.5572	0.5866	0.7007	0.6540	0.6413	0.6456	0.6465	0.6457	0.6449	0.6438
10e-2	0.5054	0.5605	0.5449	0.6980	**0.7159**	0.7006	0.6971	0.7075	0.6905	0.6766	0.7017	0.7106
10e-1	0.5036	0.5025	0.5036	0.5017	0.5070	0.5140	0.5138	0.5216	0.5184	0.5140	0.5147	0.5195
1	0.3540	0.3162	0.3118	0.3022	0.3340	0.3467	0.3500	0.3545	0.3517	0.3456	0.3630	0.3538
10e1	0.2511	0.2533	0.2691	0.2531	0.2482	0.2995	0.3116	0.3147	0.3098	0.3054	0.3077	0.3213
10e2	0.2104	0.2058	0.2204	0.2286	0.2035	0.2682	0.2794	0.2775	0.2786	0.2834	0.2765	0.2647
10e3	0.1872	0.1861	0.1964	0.1803	0.1866	0.2560	0.2657	0.2775	0.2550	0.2749	0.2682	0.2654
10e4	0.1727	0.1741	0.1636	0.1687	0.1650	0.2231	0.2297	0.2329	0.2270	0.2272	0.2208	0.2359
10e5	0.1425	0.1313	0.1294	0.1229	0.1282	0.1500	0.1597	0.1594	0.1544	0.1574	0.1565	0.1541
10e6	0.1386	0.1319	0.1385	0.1274	0.1313	0.1360	0.1375	0.1365	0.1325	0.1314	0.1391	0.1392

Table 16 Means of ARI by LSS-FTC-NTR on the subject 1+5% noise using different λ_1 and λ_2, while fixing *m*=2

λ_2 \ λ_1	0	10e-4	10e-3	10e-2	10e-1	1	10e1	10e2	10e3	10e4	10e5	10e6
0	0.7254	0.7452	0.7332	0.7778	0.7948	0.7880	0.7879	0.7876	0.7997	0.7905	0.7989	0.7845
10e-4	0.7534	0.7997	0.7590	0.8253	0.8656	0.8506	0.8670	0.8615	0.8579	0.8600	0.8460	0.8406
10e-3	0.7618	0.7943	0.8057	0.8553	0.8715	0.8733	0.8780	0.8717	0.8767	0.8755	0.8781	0.8769
10e-2	0.7586	0.8079	0.8055	0.8616	**0.8906**	0.8840	0.8840	0.8878	0.8837	0.8871	0.8849	0.8838
10e-1	0.6952	0.6842	0.6858	0.6788	0.7073	0.7172	0.7190	0.7139	0.7167	0.7215	0.7187	0.7135
1	0.6621	0.6716	0.6685	0.6733	0.6787	0.6788	0.6785	0.6797	0.6742	0.6813	0.6740	0.6748
10e1	0.6514	0.6555	0.6601	0.6612	0.6671	0.6615	0.6668	0.6665	0.6668	0.6647	0.6613	0.6649
10e2	0.6599	0.6637	0.6743	0.6744	0.6702	0.6724	0.6732	0.6746	0.6737	0.6720	0.6789	0.6771
10e3	0.6621	0.6625	0.6721	0.6583	0.6612	0.6633	0.6613	0.6650	0.6696	0.6696	0.6694	0.6666
10e4	0.6567	0.6497	0.6595	0.6536	0.6576	0.6562	0.6581	0.6597	0.6537	0.6537	0.6561	0.6512
10e5	0.6495	0.6397	0.6462	0.6357	0.6378	0.6446	0.6451	0.6448	0.6408	0.6426	0.6412	0.6388
10e6	0.6456	0.6403	0.6553	0.6402	0.6409	0.6506	0.6429	0.6419	0.6489	0.6466	0.6438	0.6439

Comments 3: In Experiments, only compared your proposed method on 5%/10%/15% noisy is insufficient. Hence, this part is a little weak. I suggest the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

authors to compare your proposed method on higher noisy scene. I would like to see the performance changes of your proposed method on different noise scenes.

Reply: We would like to thank you for your comment. In Experiments, we randomly select 20 brain CT images as the original target domain data, and the rest 236 brain CT images as source domain data. We consider the application of LSS-FTC-NTR in the scenario of target images polluted by noise. We add target images corrupted by 20%, 25% and 30% Gaussian noise. The mean and standard deviation of NMI and ARI for all compared clustering methods are displayed in Tables 4-6 and 10-12, respectively. To better observe the behavior of all algorithms, Figs. 4-9 graphically shows the segmentation results of all comparison methods obtained on subject1 with different noise. Similar to the results in the Tables 1-12, LSS-FTC-NTR obtains the best segmentation results for distinguishing the bone, water and soft issues. The boundaries between different organizations are smooth and obvious are relatively clearer than the other three methods.

Table 4

NMI performance of all comparison methods on 20% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0246	0.0289	0.5095	0.1908	0.4782	0.0069	**0.6000**	0.0055
Subject2	0.3092	0.1299	0.5796	0.1631	0.5596	0.0082	**0.6348**	0.0128
Subject3	0.2597	0.0992	0.5556	0.1588	0.5029	0.0108	**0.5926**	0.0051
Subject4	0.0102	0.0061	0.5136	0.2026	0.4778	0.0066	**0.5254**	0.0133
Subject5	0.3071	0.0852	0.5654	0.1471	0.5333	0.0119	**0.6237**	0.0084
Subject6	0.2998	0.0697	0.5662	0.1376	0.5247	0.0176	**0.5875**	0.0099
Subject7	0.0994	0.1742	0.5298	0.2377	0.4804	0.0088	**0.5719**	0.0063
Subject8	0.2557	0.1427	0.5683	0.1438	0.5426	0.0160	**0.6023**	0.0114
Subject9	0.3098	0.0342	0.5852	0.1508	0.5366	0.0089	**0.6433**	0.0062
Subject10	0.0966	0.2007	0.5788	0.2565	0.5309	0.0066	**0.5991**	0.0080
Subject11	0.1196	0.2075	0.5818	0.2101	0.5525	0.0080	**0.6516**	0.0129
Subject12	0.1530	0.1197	0.5707	0.1767	0.5466	0.0167	**0.5869**	0.0122

Table 5

NMI performance of all comparison methods on 25% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0107	0.0058	0.5078	0.2154	0.4969	0.0043	**0.5469**	0.0104
Subject2	0.1708	0.1001	0.5792	0.1741	0.5842	0.0076	**0.6572**	0.0182
Subject3	0.2460	0.0981	0.5503	0.1685	0.5232	0.0145	**0.5940**	0.0093
Subject4	0.0121	0.0081	0.5177	0.2390	0.5045	0.0112	**0.5717**	0.0197
Subject5	0.1503	0.1169	0.5621	0.1599	0.5580	0.0122	**0.6090**	0.0083
Subject6	0.2199	0.0660	0.5662	0.1533	0.5434	0.0311	**0.5935**	0.0142
Subject7	0.0284	0.0285	0.5331	0.2528	0.5156	0.0137	**0.5677**	0.0095

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Subject8	0.1320	0.1375	0.5705	0.1577	0.5547	0.0156	**0.6272**	0.0136
Subject9	0.2594	0.1054	0.5859	0.1542	0.5682	0.0124	**0.6437**	0.0127
Subject10	0.1980	0.0293	0.5763	0.2203	0.5488	0.0073	**0.6416**	0.0157
Subject11	0.2156	0.2697	0.5830	0.2349	0.5745	0.0184	**0.6732**	0.0117
Subject12	0.3935	0.0908	0.5724	0.1827	0.5617	0.0122	**0.6386**	0.0078

Table 6

NMI performance of all comparison methods on 30% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0120	0.0093	0.5086	0.2028	0.5569	0.0124	**0.6567**	0.0084
Subject2	0.2553	0.1770	0.5833	0.1497	0.6333	0.0113	**0.7439**	0.0109
Subject3	0.2010	0.0804	0.5526	0.1537	0.5872	0.0206	**0.6602**	0.0110
Subject4	0.0891	0.1752	0.5143	0.2144	0.5664	0.0050	**0.6720**	0.0114
Subject5	0.1820	0.1376	0.5663	0.1463	0.6172	0.0193	**0.7080**	0.0028
Subject6	0.2316	0.1094	0.5695	0.1928	0.6062	0.0120	**0.6872**	0.0086
Subject7	0.1080	0.1686	0.5316	0.2091	0.5769	0.0172	**0.6966**	0.0054
Subject8	0.1755	0.1045	0.5696	0.1777	0.6231	0.0097	**0.7037**	0.0103
Subject9	0.2584	0.0893	0.5913	0.1901	0.6295	0.0216	**0.7356**	0.0200
Subject10	0.0113	0.0135	0.5751	0.2901	0.6247	0.0198	**0.7441**	0.0115
Subject11	0.2121	0.2838	0.5827	0.2652	0.6343	0.0212	**0.7491**	0.0042
Subject12	0.1547	0.1170	0.5702	0.1941	0.6226	0.0045	**0.7256**	0.0095

Table 10

ARI performance of all comparison methods on 20% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0019	0.0080	0.3400	0.1486	0.3241	0.0059	**0.6624**	0.0098
Subject2	0.1754	0.1208	0.3933	0.1522	0.3859	0.0059	**0.6005**	0.0116
Subject3	0.1443	0.0831	0.4147	0.1533	0.3742	0.0102	**0.5570**	0.0066
Subject4	0.0015	0.0060	0.3464	0.1713	0.3270	0.0090	**0.5008**	0.0139
Subject5	0.1679	0.0931	0.3883	0.1370	0.3730	0.0056	**0.6152**	0.0068
Subject6	0.1701	0.0682	0.4176	0.1304	0.3838	0.0091	**0.5775**	0.0079
Subject7	0.0625	0.1226	0.3602	0.1912	0.3309	0.0064	**0.4983**	0.0083
Subject8	0.1520	0.1066	0.3913	0.1339	0.3809	0.0088	**0.5055**	0.0102
Subject9	0.1619	0.0404	0.4164	0.1424	0.3871	0.0056	**0.5725**	0.0053
Subject10	0.0620	0.1394	0.3812	0.1927	0.3571	0.0027	**0.5494**	0.0065
Subject11	0.0713	0.1457	0.3838	0.1649	0.3695	0.0065	**0.5819**	0.0122
Subject12	0.0657	0.0780	0.3746	0.1617	0.3641	0.0048	**0.4376**	0.0106

Table 11

ARI performance of all comparison methods on 25% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0023	0.0020	0.3409	0.1786	0.3419	0.0056	**0.5128**	0.0139
Subject2	0.0687	0.0687	0.3948	0.1626	0.4354	0.0132	**0.6609**	0.0156
Subject3	0.1447	0.0664	0.4126	0.1659	0.4043	0.0113	**0.5219**	0.0102
Subject4	0.0125	0.0047	0.3484	0.1856	0.3738	0.0138	**0.5349**	0.0204
Subject5	0.0701	0.0553	0.3862	0.1545	0.4182	0.0235	**0.6445**	0.0077
Subject6	0.1147	0.0392	0.4186	0.1502	0.4153	0.0305	**0.5271**	0.0151
Subject7	0.0076	0.0109	0.3611	0.1885	0.3654	0.0088	**0.5132**	0.0116
Subject8	0.0656	0.0666	0.3914	0.1499	0.4143	0.0176	**0.5902**	0.0148
Subject9	0.1248	0.0841	0.4173	0.1462	0.4233	0.0152	**0.6459**	0.0121
Subject10	0.1157	0.0035	0.3804	0.1958	0.3796	0.0046	**0.5953**	0.0165
Subject11	0.1382	0.1889	0.3842	0.1686	0.4277	0.0200	**0.6537**	0.0123
Subject12	0.2421	0.0869	0.3739	0.1749	0.4058	0.0177	**0.6031**	0.0091

Table 12

ARI performance of all comparison methods on 30% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0008	0.0052	0.3401	0.1505	0.5432	0.0294	**0.7743**	0.0062
Subject2	0.1663	0.1180	0.3941	0.1505	0.5887	0.0309	**0.8122**	0.0136
Subject3	0.0947	0.0389	0.4132	0.1584	0.5700	0.0397	**0.7414**	0.0125
Subject4	0.0601	0.1242	0.3471	0.1760	0.5526	0.0094	**0.7876**	0.0146
Subject5	0.0867	0.0967	0.3886	0.1351	0.5858	0.0488	**0.7966**	0.0044
Subject6	0.1193	0.0774	0.4173	0.1744	0.5833	0.0131	**0.7834**	0.0053
Subject7	0.0675	0.1173	0.3611	0.1547	0.5637	0.0408	**0.7874**	0.0079
Subject8	0.0684	0.0490	0.3911	0.1728	0.6018	0.0277	**0.7363**	0.0144
Subject9	0.1227	0.0787	0.4184	0.1823	0.6060	0.0214	**0.8233**	0.0176
Subject10	0.0047	0.0076	0.3800	0.2041	0.5840	0.0421	**0.8140**	0.0134
Subject11	0.1420	0.1947	0.3839	0.2030	0.6010	0.0318	**0.8226**	0.0052
Subject12	0.0588	0.0640	0.3745	0.1792	0.5793	0.0106	**0.7996**	0.0093

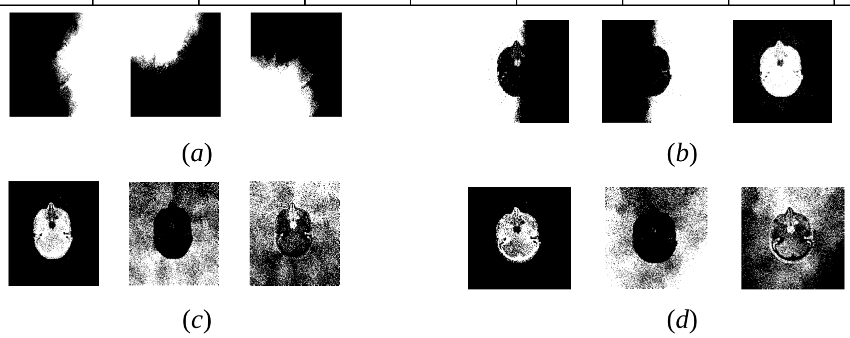

Fig.7 Clustering segmentations on subject1+20% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

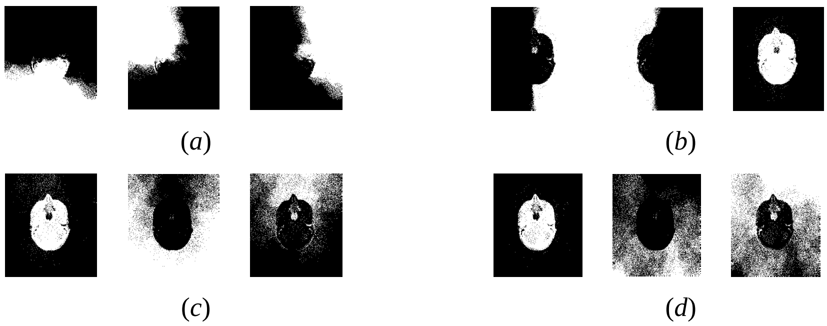

Fig.8 Clustering segmentations on subject1+25% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

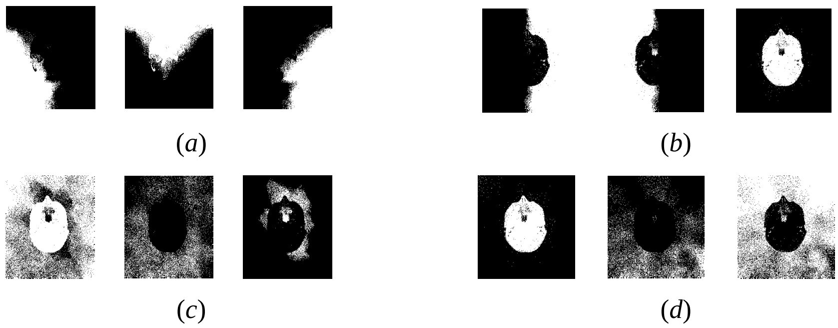

Fig.9 Clustering segmentations on subject1+30% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

Comments 4: About the speed of the compared approaches, does the new proposed approach affect the speed?

Reply: Thank you for your comments. The time complexity of LSS-FTC-NTR is dependent on matrix operation and the number of iterations of gradient descent algorithm. The time complexity of matrix inversion is the cubic of the scale of training set. Thus, LSS-FTC-NTR is slower than its baseliner method FCM. Thus, how to speed up LSS-FTC-NTR is worthy to be studied in the future. We aim to propose a fast version of LSS-FTC-NTR to handle with large scale brain CT image segmentation tasks. In response to your comments, we add this issue in the conclusion in the revised manuscript. (***Ref to the conclusion section***):

Future work will extend our algorithm to other medical image segmentation applications. ...In addition, how to speed up LSS-FTC-NTR is worthy to be studied in the future.

Comments 5: In the Conclusion section, the limitations of the proposed measure are scarcely discussed.

Reply: Thank you for your comments. We have added more discussions about the limitations of the proposed method in the conclusion in the revised manuscript. (***Ref to the conclusion section***):

Future work will extend our algorithm to other medical image segmentation applications. We will extend the framework so as to apply various clustering algorithms in order to obtain more satisfactory medical image segmentation results. We will also study how many images in the source domain can be considered sufficient, and how to select the important images to further improve the transfer. In addition, how to speed up LSS-FTC-NTR is worthy to be studied in the future.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Comments 6: The proposed method is very related to co-clustering, some latest important references should be included.

Reply: Thank you for your comments. Co-clustering, also called simultaneously clustering, performs clustering on both rows and columns of data samples, which aims to exploit the clear duality between rows and columns of a contingency table. The difference between transfer learning and co-clustering is that transfer clustering promotes clustering performance in the current domain by leveraging knowledge from the other related domains, and transfer learning does not need the hypothesis that the given data samples satisfy the independent and identically distributed (IID) of training samples and testing samples. In practical, a model with very limited brain CT images will not be able to train for image segmentation. We try to use a large amount of samples of brain CT images collected from different hospitals as auxiliary knowledge. Thus, we prefer transfer learning in this study.

In our revised manuscript, we discuss this issue in the subsection 2.2. **(*Ref to: the last paragraph of page 3*)**:

Currently, when the training data is not enough to represent the current domain, transfer learning, multi-task learning and co-clustering are three effective techniques that can enhance the clustering performance in the current domain. Multi-task learning performs multiple learning tasks together through by sharing certain knowledge among all tasks [30, 31]. Co-clustering performs clustering on both rows and columns of data samples to exploit the clear duality between rows and columns of a contingency table [32]. Transfer learning clustering enhances the clustering performance in the new domain by leveraging useful knowledge from different but related domains. Many researches show that the transfer clustering methods have better learning ability to obtain an effective model with the idea of transfer learning [28, 29, 33]. In real applications, due to the existences of noise and field offset etc, the insufficient medical images are inadequate to complete image segmentation. Therefore, we think transfer learning clustering is an effect technology to promote the segmentation of insufficient and noisy medical image in the new domain.

Comments 7: The paper should be further proofread carefully. Some minor grammatical errors are in it.

Reply: Thanks a lot for your reminder. We have carefully proofed the manuscript to correct grammatical errors. We also have asked a technical writer to polish the manuscript. We believe the quality of the revised manuscript has been improved significantly.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Response to Reviewer 5

Comments 1: There is currently some weaknesses w.r.t. to other criteria which should be addressed in a minor revision round: For the comparison with related work, is the number of clusters initially assumed to be known for all algorithms used in the comparison?

Reply: Thank you. In the experiments, we perform our experiments on ultrashort echo time (UTE) and modified Dixon brain image datasets. All CT images with corresponding manual segmentation are segmented into three classes: bone, water and soft issues. Thus, for the comparison with related work, the number of clusters is manually set to be three.

Comments 2: Author should give a flowchart about the proposed algorithm, which increases our understanding. As a peer, my knowledge is very narrow and needs a large framework to understand.

Reply: Thank you for your comments. In order to follow your suggestions, we have made the following changes in the revised manuscript:

1) We have added up the motivation of LSS-FTC-NTR in Section 1 (***Ref to: the second paragraph of page 3***):

The motivation of LSS-FTC-NTR is shown in Fig.1. Two cluster centers presented as black triangle and circle have positive transfer influence to the clustering in the target domain, while the cluster center presented as Black Square has negative influence to the clustering in the target domain. LSS-FTC-NTR will automatically resist black square participating in the clustering in the target domain by using the negative-transfer-resistance strategy.

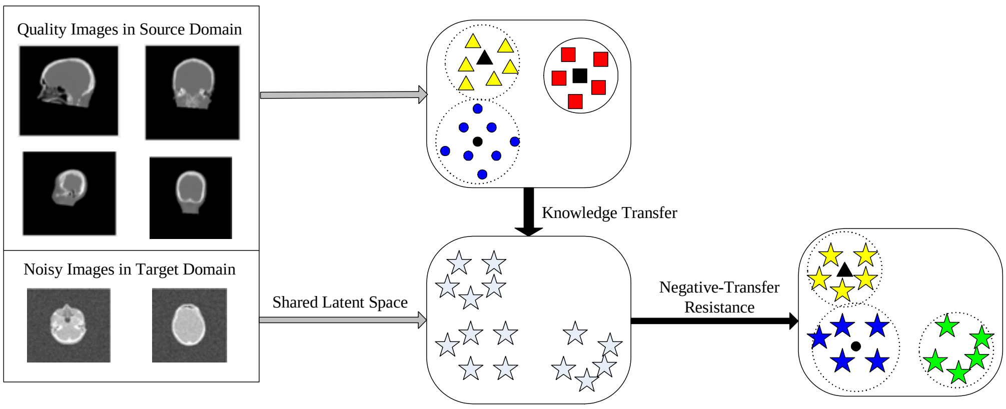

Fig.1 The motivation of LSS-FTC-NTR

2) We have added up the novelty of LSS-FTC-NTR in Section 1 (***Ref to: the second paragraph of page 3***):

The novelty of this study is as follows.

1) We formulate the problem of insufficient and noisy medical image segmentation as a model of transfer clustering task. To the best of our knowledge, our study is the first attempt to address this issue.

2) The negative-transfer-resistance mechanism is proposed to identify and resist negative source transfer knowledge.

3) The MMD is introduced into LSS-FTC-NTR to unify the representation of image data of different

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

domains in the shared latent space, which helps transferring knowledge across domains.

4) Clustering centers based transfer matching scheme is used to deal with the inconsistency problem of clustering numbers between source and target domains, so that more robust and cluster performance can be promoted.

3) To clearly describe the algorithm of LSS-FTC-NTR, we have re-written some steps of Algorithm1 (***Ref to: the second paragraph of page 3***):

Algorithm 1: LSS-FTC-NTR model	
Initialize	Set the maximum number of iterations t_{max}, the fuzzy index m, the regularization parameters λ_1 and λ_2, and the learning rate η.
Repeat:	
Exacting transfer knowledge form the source domain;	
Perform soft-partition clustering methods in the source domain, such as FCM, and obtain the cluster centers of data in the source domain;	
$t = t+1$;	
Initialize the clustering centers of data in the target domain;	
Compute the weight of transfer knowledge S_{jh} using Eq. (10);	
Fix $\mathbf{U}(t)$ and $\mathbf{\Theta}(t)$, obtain $\tilde{\mathbf{V}}^{TD}(t)$ using Eq. (12);	
Fix $\tilde{\mathbf{V}}^{TD}(t)$ and $\mathbf{\Theta}(t)$, obtain $\mathbf{U}(t)$ using Eq. (14);	
Fix $\mathbf{U}(t)$ and $\tilde{\mathbf{V}}^{TD}(t)$, obtain $\mathbf{\Theta}(t)$ using Eq. (18) and Eq.(19);	
Compute $J(t)$ using Eq. (9);	
Until $\|J(t)-J(t-1)\| \le \delta$ or $t \ge t_{max}$;	

Comments 3: Are there any specific restrictions for \lambda_{\a} and \lambda_{\b} in Eq.(9)?

Reply: Thank you. In the experiments, we select λ_1 and λ_2 in a given search grid.

In our revised manuscript, in response to both your comment and another referee's suggestions, we discuss the performance of LSS-FTC-NTR using different parameters. (***Ref to: the second paragraph of page 3***):

Tables 15-16 show the means of NMI and ARI on the subject using different λ_1 and λ_2, while fixing the parameter m=2. We can see that

1) LSS-FTC-NTR is sensitive to parameters λ_1 and λ_2. Different λ_1 and λ_2 lend to different cluster performance of LSS-FTC-NTR in terms of NMI and ARI. It can be found that in most situations when the value of NMI is better, the value of ARI is also better. Thus, it is feasible to use NMI and ARI as performance criterions to determine the suitable parameters.

2) Fixed the value of m, LSS-FTC-NTR obtains the worst NMI and ARI when $\lambda_1 = 0$ and $\lambda_2 = 0$. The clustering performance of LSS-FTC-NTR is improved when λ_1 and λ_2 are not equal to 0. Since when $\lambda_1 = 0$ and $\lambda_2 = 0$ LSS-FTC-NTR is degenerated to the classical FCM clustering.

3) We can find that when the value of λ_1 is large, LSS-FTC-NTR obtains the satisfactory performance in terms of NMI and ARI. This further demonstrates that the proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

negative-transfer-resistance mechanism has played an effective role. Thus, in the subsequent experiments, we can reduce the search grid of λ_1 *in the range* $\{10e1, 10e1, ..., 10e6\}$. *We can't find the rule to select parameter* λ_2. *We think it is reasonable to select optimal* λ_2 *within the search grid. The range* $\lambda_2 \in \{10e\text{-}4, 10e\text{-}3, ..., 10e6\}$ *is appropriate.*

Table 15 Means of NMI by LSS-FTC-NTR on the subject1+5% noise using different λ_1 and λ_2, while fixing *m*=2

λ_2 \ λ_1	0	10e-4	10e-3	10e-2	10e-1	1	10e1	10e2	10e3	10e4	10e5	10e6
0	0.4801	0.5006	0.5181	0.5501	0.6011	0.6250	0.6091	0.6375	0.6788	0.6397	0.6427	0.6378
10e-4	0.4915	0.5326	0.5592	0.5662	0.6109	0.6161	0.6469	0.6499	0.6468	0.6493	0.6431	0.6425
10e-3	0.5094	0.5436	0.5572	0.5866	0.7007	0.6540	0.6413	0.6456	0.6465	0.6457	0.6449	0.6438
10e-2	0.5054	0.5605	0.5449	0.6980	**0.7159**	0.7006	0.6971	0.7075	0.6905	0.6766	0.7017	0.7106
10e-1	0.5036	0.5025	0.5036	0.5017	0.5070	0.5140	0.5138	0.5216	0.5184	0.5140	0.5147	0.5195
1	0.3540	0.3162	0.3118	0.3022	0.3340	0.3467	0.3500	0.3545	0.3517	0.3456	0.3630	0.3538
10e1	0.2511	0.2533	0.2691	0.2531	0.2482	0.2995	0.3116	0.3147	0.3098	0.3054	0.3077	0.3213
10e2	0.2104	0.2058	0.2204	0.2286	0.2035	0.2682	0.2794	0.2775	0.2786	0.2834	0.2765	0.2647
10e3	0.1872	0.1861	0.1964	0.1803	0.1866	0.2560	0.2657	0.2775	0.2550	0.2749	0.2682	0.2654
10e4	0.1727	0.1741	0.1636	0.1687	0.1650	0.2231	0.2297	0.2329	0.2270	0.2272	0.2208	0.2359
10e5	0.1425	0.1313	0.1294	0.1229	0.1282	0.1500	0.1597	0.1594	0.1544	0.1574	0.1565	0.1541
10e6	0.1386	0.1319	0.1385	0.1274	0.1313	0.1360	0.1375	0.1365	0.1325	0.1314	0.1391	0.1392

Table 16 Means of ARI by LSS-FTC-NTR on the subject 1+5% noise using different λ_1 and λ_2, while fixing *m*=2

λ_2 \ λ_1	0	10e-4	10e-3	10e-2	10e-1	1	10e1	10e2	10e3	10e4	10e5	10e6
0	0.7254	0.7452	0.7332	0.7778	0.7948	0.7880	0.7879	0.7876	0.7997	0.7905	0.7989	0.7845
10e-4	0.7534	0.7997	0.7590	0.8253	0.8656	0.8506	0.8670	0.8615	0.8579	0.8600	0.8460	0.8406
10e-3	0.7618	0.7943	0.8057	0.8553	0.8715	0.8733	0.8780	0.8717	0.8767	0.8755	0.8781	0.8769
10e-2	0.7586	0.8079	0.8055	0.8616	**0.8906**	0.8840	0.8840	0.8878	0.8837	0.8871	0.8849	0.8838
10e-1	0.6952	0.6842	0.6858	0.6788	0.7073	0.7172	0.7190	0.7139	0.7167	0.7215	0.7187	0.7135
1	0.6621	0.6716	0.6685	0.6733	0.6787	0.6788	0.6785	0.6797	0.6742	0.6813	0.6740	0.6748
10e1	0.6514	0.6555	0.6601	0.6612	0.6671	0.6615	0.6668	0.6665	0.6668	0.6647	0.6613	0.6649
10e2	0.6599	0.6637	0.6743	0.6744	0.6702	0.6724	0.6732	0.6746	0.6737	0.6720	0.6789	0.6771
10e3	0.6621	0.6625	0.6721	0.6583	0.6612	0.6633	0.6613	0.6650	0.6696	0.6696	0.6694	0.6666
10e4	0.6567	0.6497	0.6595	0.6536	0.6576	0.6562	0.6581	0.6597	0.6537	0.6537	0.6561	0.6512
10e5	0.6495	0.6397	0.6462	0.6357	0.6378	0.6446	0.6451	0.6448	0.6408	0.6426	0.6412	0.6388
10e6	0.6456	0.6403	0.6553	0.6402	0.6409	0.6506	0.6429	0.6419	0.6489	0.6466	0.6438	0.6439

Comments 4: Please let us know what is the most important factor to affect the proposed method?

Reply: Thank you for your comments. There are three free parameters in LSS-FTC-NTR, including fuzzy index *m*, parameters λ_1 and λ_2. The fuzzy index m is fixed to be 2, and λ_1 and λ_2 is selected in a given search grid. In our revised manuscript, in response to both your comment and another referee's suggestions, we discuss the performance of LSS-FTC-NTR using different parameters. The experimental results are shown in Tables 15-16, respectively. We can find that LSS-FTC-NTR is sensitive to parameters λ_1 and λ_2. Different λ_1 and λ_2 lend to different cluster performance of LSS-FTC-NTR in terms of NMI and ARI. It can be found that in most situations when the value of NMI is better, the value of ARI is also better. We can find that when the value of λ_1 is large, LSS-FTC-NTR obtains the satisfactory performance in terms of NMI and ARI. This further demonstrates that the proposed negative-transfer-resistance mechanism has played an effective role. Thus, in the subsequent experiments, we can reduce the search grid of λ_1 in the range $\{10e1, 10e1, \ldots, 10e6\}$. We can't find the rule to select parameter λ_2. We think it is reasonable to select optimal λ_2 within the search grid. The range $\lambda_2 \in \{10e\text{-}4, 10e\text{-}3, \ldots, 10e6\}$ is appropriate.

A Novel Negative-Transfer-Resistant Fuzzy Clustering Model with a Shared Cross-Domain Transfer Latent Space and its Application to Brain CT Image Segmentation

Yizhang Jiang, *Senior Member, IEEE*, Xiaoqing Gu, Dongrui Wu, *Senior Member, IEEE*, Wenlong Hang, Jing Xue, Shi Qiu，Chin-Teng Lin, *Fellow, IEEE*

Abstract—Traditional clustering algorithms for medical image segmentation can only achieve satisfactory clustering performance under relatively ideal conditions, in which there is adequate data from the same distribution, and the data is seldom disturbed by noise or outliers. However, a sufficient amount of medical images with representative manual labels are often not available, because medical images are frequently acquired with different scanners (or different scan protocols) or polluted by various noises. Transfer learning improves learning in the target domain by leveraging knowledge from related domains. Given some target data, the performance of transfer learning is determined by the degree of relevance between the source and target domains. To achieve positive transfer and avoid negative transfer, a negative-transfer-resistant mechanism is proposed by computing the weight of transferred knowledge. Extracting a negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space (called LSS-FTC-NTR) is proposed by integrating negative-transfer-resistant and maximum mean discrepancy (MMD) into the framework of fuzzy c-means clustering. Experimental results show that the proposed LSS-FTC-NTR model outperformed several traditional non-transfer and related transfer clustering algorithms.

Index Terms—medical image segmentation, fuzzy clustering, transfer learning, negative transfer

1 Introduction

WITH the development of electronic information and computer technology, medical imaging and image processing technology have developed rapidly. Medical imaging equipments collect images for a short time, and are less affected by external factors. Today, medical imaging technology has become a powerful tool and core technology for modern clinical diagnosis and treatment. Commonly used medical imaging techniques include Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Computed Radiography (CR), Ultrasound, and so on. CT scans produce clearer images than conventional x-ray for internal organs, bone and soft tissues. MRI scans furnish greater clearness and higher resolution than CT scans with lower resolution [1]. Medical image segmentation is a basic and important step in medical image processing and analysis. It is also the basis of medical image registration, medical image information fusion and 3D visualization. In the current clinical practice, manual segmentation based on visual recognition and empirical judgment by doctors is still the most typical and common segmentation method. However, manual segmentation is tedious, time consuming and subjective. For example, in the Isointense Infant Brain Segmentation Challenge (ISEG2017), manual segmentation of each brain MRI scan took an average of one week for neuroradiologists [2]. Furthermore, the differences in physician experience and uncertain factors such as visual fatigue will affect the correct analysis of segmentation results.

With the rapid growth of the image processing technology, many automatic image processing techniques have appeared in recent years [3-5]. Image segmentation approaches can be broadly classified into four categories: graph based methods, classification methods, deep learning methods and clustering methods [6]. A medical image in graph based image segmentation is presented as a weighted undirected

- *Y. Jiang is with the Jiangsu Key Laboratory of Media Design and Software Technology, and is also with the School of Digital Media, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, China. E-mail: yzjiang@jiangnan.edu.cn.*
- *X. Gu is with the School of Information Science and Engineering, Changzhou University, Changzhou, Jiangsu 213164, China. E-mail: czxqgu@163.com.*
- *D. Wu is with the Key Laboratory of the Ministry of Education for Image Processing and Intelligent Control, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China E-mail: drwu09@gmail.com.*
- *W. Hang is with the School of Computer Science and Technology, Nanjing Tech University, Jiangsu, Nanjing, China. E-mail: hangwenlongnj@sina.com.*
- *J. Xue is with the Department of Nephrology, the Affiliated Wuxi People's Hospital of Nanjing Medical University, 299 Qingyang Rd, Wuxi, Jiangsu 214023, China. E-mail: 475809491@qq.com.*
- *S. Qiu is with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, 710119, China. E-mail: qiushi215@163.com.*
- *C.-T. Ling is with the Centre of Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia. E-mail: ChinTeng.Lin@uts.edu.au.*

Corresponding authors: Xiaoqing Gu and Yizhang Jiang

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

graph [7]. Each pixel or region in the image is treated as a vertex of a graph, and the set of edges can be connected by adjacent pixels or two adjacent regions. Then the image is divided into several parts according to the relationship of the adjacent pixels. The second category, also called supervised methods, use labeled segmented images to extract features and train a segmentation model, such as *k*-nearest neighbor (KNN) [8], neural network [9], support vector machine (SVM) [10], and so on. A major drawback of classification method is that it requires sufficient labeled training images. But in the areas of medical imaging, it is relatively easy and inexpensive to obtain a large amount of unlabeled data [11]. Deep learning learns the feature representation of tissue contour based on deep convolutional neural networks [12, 13]. Deep learning methods have successfully applied for medical image segmentation in recent years. However, deep learning methods usually need a large number of training dataset and special hardware devices. It is known that the medical image segmentation problem can be considered as classifying the pixels of images into homogeneous regions. This process can be viewed as clustering problem. Clustering method, as an unsupervised machine learning approach, is the process of grouping a set of data points into subsets so that data points in the same subset are similar (according to some criteria). Widely used clustering methods include expectation-maximization, spectral clustering, fuzzy clustering, and so on. In the last decade, clustering-based algorithms have attracted great interest in the segmentation of medical images [14]. Portela et al [15] proposed clustering based semi-supervised classification for brain image segmentation, which used *K*-means clustering as initial processing to select brain slices. Ortiz et al [16] improved brain image segmentation by using self-organizing maps (SOMs) based voxel clustering to extract features, and then used entropy-gradient clustering to segment brain images. Based on the idea of multiobjective optimization, Saha et al [17] proposed semi-supervised clustering in the intensity space for medical image segmentation. Abdel-Maksouda et al [18] combined *K* means and FCM clustering to propose a hybrid clustering approach for tumour segmentation from brain image.

In order for clustering based segmentation methods to perform well, the training medical image data needs to be representative of the target data. However, medical images are often collected with different scanners and scanning parameters, medical images may have large differences in image quality due to machine performance or scanning technology, such as varying degrees of rotation, noise, etc. The requirement of training and target data under the same distribution prevents the use of clustering algorithms in larger research and clinical practice. Since the above scenarios exist in a large number of real-world environments, this leads to unsatisfactory segmentation results and the risk of algorithm failure. To solve this problem, researchers have introduced the idea of transfer learning into clustering algorithms [19, 20]. With the help of some knowledge of auxiliary domain (called source domain), transfer learning handles the cases where the distribution, feature space or tasks are different between source domain and test domain (called target domain). In transfer learning, the auxiliary knowledge from the source domain involves the data sample, feature representations, parameters and relationships [21-23]. The knowledge is usually obtained from certain precise procedures and reliable theory through some specific perspectives. Jiang et al [24] proposed a transfer spectral clustering approach, which uses both data manifold and feature manifold between related clustering tasks. Deng et al [25] proposed a transfer prototype-based fuzzy clustering approach, which incorporated prototype knowledge induced from source domain to implement the clustering in the target domain. Qian et al [20] proposed a cross-domain maximum entropy clustering approach, which utilized the auxiliary knowledge from cluster centers and fuzzy memberships belonging to source data. However, these algorithms have a common assumption that source domain and target domain must have the same number of clusters. Moreover, most existing transfer clustering methods are not developed for noisy scenarios. Thus, these methods may be not suitable for noisy medical image segmentation.

Since the medical images of different domains may have variations due to changes caused by noise, field offset and bias field, in this paper, we study the problem of medical image segmentation in a noisy scenario by transferring medical images collected from related scenarios. We consider the new noisy scenario as the target domain and the existing medical image dataset from related scenario elsewhere as the source domain, and then use the learning on clean images of source data to improve the clustering in target data. To improve the transfer learning performance, we consider learning the negative-transfer-resistant mechanism, so that the influence of positive transfer knowledge is reinforced and the influence of negative transfer knowledge is reduced or even eliminated. Meanwhile, we think medical images in different scenarios share certain common representations such as bone and soft tissue, and the shared representations could be preserved in a shared space. Inspired of maximum mean discrepancy (MMD)[26], we learn the shared latent space for source and target domains such that the distributions in different domains are close to each other. We investigate transferring ability of each cluster belonging to source domain in the shared latent space for medical image segmentation modeling. We use the clustering centers in the source domain as the transfer knowledge, regardless of whether the number of clusters in the source domain is the same as that in the target domain. With the above ideas, we propose a negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space (called

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

LSS-FTC-NTR). The motivation of LSS-FTC-NTR is shown in Fig.1. Two cluster centers presented as black triangle and circle have positive transfer influence to the clustering in the target domain, while the cluster center presented as Black Square has negative influence to the clustering in the target domain. LSS-FTC-NTR will automatically resist black square participating in the clustering in the target domain by using the negative-transfer-resistance strategy. We evaluate the proposed LSS-FTC-NTR method on real world datasets, and compare with several non-transfer and related transfer clustering methods. The results on real world brain CT dataset demonstrate that LSS-FTC-NTR is more robust than the comparison methods.

The novelty of this study is as follows. 1) We formulate the problem of insufficient and noisy medical image segmentation as a model of transfer clustering task. To the best of our knowledge, our study is the first attempt to address this issue. 2) The negative-transfer-resistance mechanism is proposed to identify and resist negative source transfer knowledge. 3) The MMD is introduced into LSS-FTC-NTR to unify the representation of image data of different domains in the shared latent space, which helps transferring knowledge across different domains. 4) Clustering centers based transfer matching scheme is used to deal with the inconsistency problem of clustering numbers between source and target domains, so that more robust cluster performance can be promoted.

The rest of this paper is organized as follows. Concepts related to FCM, transfer learning and MMD are reviewed in Section II. In Section III, the negative-transfer-resistant mechanism and new LSS-FTC-NTR algorithm is introduced. Its parameter learning based on iteratively optimization strategy is then presented accordingly. The experimental results in real-world brain CT image datasets are reported in Section IV. Conclusions are given in the last section.

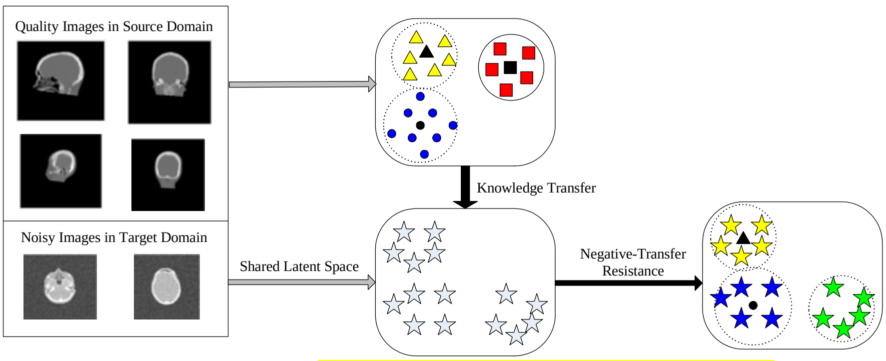

Fig.1 The motivation of LSS-FTC-NTR

2 Related Work

2.1 Conventional FCM for Image Segmentation

Fuzzy *C*-means (FCM) clustering [27] is one of most commonly used fuzzy clustering algorithms. FCM allows data points to belong to more than one cluster defined by a membership matrix. Let $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N\}$ be a given dataset where d and N are data dimension and capacity, respectively. Suppose there exist C clusters in $\mathbf{X}$, FCM derives the following objection function:

$$\min_{\mathbf{U},\mathbf{V}} J=\sum_{i=1}^{N}\sum_{j=1}^{C}\mu_{ij}^{m}\left\|\mathbf{x}_i-\mathbf{v}_j\right\|^2,$$

$$\text{s.t. } 0\le\mu_{ij}\le 1, \sum_{j=1}^{C}\mu_{ij}=1, \tag{1}$$

where $\mathbf{U}=[\mu_{ij}]_{N\times C}$ denotes the fuzzy membership matrix, and $\mathbf{V}=[\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_C]^T$ denotes the clustering center matrix. m (m>1) denotes the fuzzy index.

FCM finds an optimal group of sets to explain the data samples into C clusters via matrixes $\mathbf{Y}$ and $\mathbf{U}$. FCM minimizes the total membership weighted distance of each sample $\mathbf{x}_i$ to the clustering center $\mathbf{v}_j$. FCM can easily optimize the objective functions by an iterative technique. Among cluster center-based clustering methods, FCM was simple, efficient and high popularity. It is widely used in transfer clustering methods.

For example, a FCM-based transfer learning was proposed in [28], which is combined with Gini-Simpson diversity index and quadratic weights on membership. A knowledge-leveraged transfer FCM (KL-TFCM) is proposed in [29], which uses three-interlinked framework of knowledge extraction, knowledge matching, and knowledge utilization to leverage source information to help the FCM clustering in the target domain. However, the above FCM-based transfer learning clustering methods are completed in the original space, and while they do not consider resisting negative transfer.

2.2 Transfer Learning and Maximum Mean Discrepancy

Currently, when the training data is not enough to represent the current domain, transfer learning, multi-task learning and co-clustering are three effective techniques that can enhance the clustering performance in the current domain. Multi-task learning performs multiple learning tasks together through by sharing certain knowledge among all tasks [30, 31]. Co-clustering performs clustering on both objects and features to exploit the clear duality between rows and columns of a contingency table [32]. Transfer learning clustering enhances the clustering performance in the new domain by leveraging useful knowledge from different but related domains. Many researches show that the transfer clustering methods have better learning ability to obtain an effective model with the idea of transfer learning [28, 29, 33]. In real applications, due to the existences of noise and field offset etc, the insufficient medical images are inadequate to complete image segmentation. Therefore, we think transfer learning clustering can be used as an effect technology to promote the segmentation of insufficient and noisy medical image in the new domain.

In transfer learning, a fundamental problem is to evaluate the distribution difference between source domain and target domain. Many criteria, like Kullback-Leibler (KL) divergence can be used for distribution estimation. But some need density estimation, are parametric and not suitable for high-dimensional data [34, 35]. In these cases, maximum mean discrepancy (MMD) as a nonparametric estimate criterion receives is widely used for comparing distributions. Let $\mathbf{X}_s=\{\mathbf{x}_{1,s},\mathbf{x}_{2,s},\ldots,\mathbf{x}_{N_s,s}\}$ and $\mathbf{X}_t=\{\mathbf{x}_{1,t},\mathbf{x}_{2,t},\ldots,\mathbf{x}_{N_t,t}\}$ denote the samples from each distribution $P_{source}(\mathbf{X}_s)$ and $P_{target}(\mathbf{X}_t)$ belonging to source domain and target domain, respectively. The MMD for comparing distributions between $P_{source}(\mathbf{X}_s)$ and $P_{target}(\mathbf{X}_t)$ is defined as

$$Dist(P_{source}(\mathbf{X}_s),P_{target}(\mathbf{X}_t))=\left\|\frac{1}{N_s}\sum_{i=1}^{N_s}f(\mathbf{x}_{i,s})-\frac{1}{N_t}\sum_{i=1}^{N_t}f(\mathbf{x}_{i,t})\right\|^2 \quad (2)$$

MMD is based on reproducing kernel Hilbert space (RKHS). Suppose $f:\mathbf{X}\to\mathbf{H}$, $\mathbf{H}$ is a universal RKHS. By inducing nonlinear mapping ϕ, function evaluation can be represented $f(\mathbf{x})=<\phi(\mathbf{x}),f>$, then equation (2) can be rewritten as

$$Dist(P_{source}(\mathbf{X}_s),P_{target}(\mathbf{X}_t))=\left\|\frac{1}{N_t}\sum_{i=1}^{N_t}\phi(\mathbf{x}_{i,t})-\frac{1}{N_s}\sum_{i=1}^{N_s}\phi(\mathbf{x}_{i,s})\right\|^2. \quad (3)$$

When the difference between source and target domains is small, the relationship between two domains is strong and the transfer knowledge can be taken full advantage of. However, when the data in the source domain is not sufficiently related, the clustering performance in the target domain may not only fail to promote, it may even actually decrease. Thus, transfer learning would resist negative transfer when source and target domains are not a good match. One strategy of resisting negative transfer is to identify and reject unhelpful knowledge from source domain. Some data selection and source selection algorithms have been proposed. The former implements some rules to select data samples to reconstruct the training set of source domain. For example, Rosenstein et al. [36] proposed a detecting negative transfer algorithm based on naive Bayes classification model. Croonenborghs et al. [37] proposed an option-based transfer in reinforcement learning algorithm to achieve a balance between positive and negative transfer. The source selection algorithms are applicable for multiple source scenarios, which select the best source domain (task) for transfer learning. An example of this strategy is Talvitie and Singh [38] proposed a Markov decision process to select the proper source task.

3 Negative-Transfer-Resistant Fuzzy Clustering Model with a Shared Cross-Domain Transfer Latent Space

The schematic diagram of LSS-FTC-NTR is shown in Fig.2. In the noisy transfer learning scenario, compared with data sample, feature representations, parameters and relationships are usually considered as being more insightful and more resistant to noise. In this study, we use the cluster centers in the source domain as auxiliary knowledge. The reason is that cluster centers are computed by certain reliable theories and rigorous procedures; such that the obtained cluster centers can well represent a cluster and all affiliated samples in a cluster.

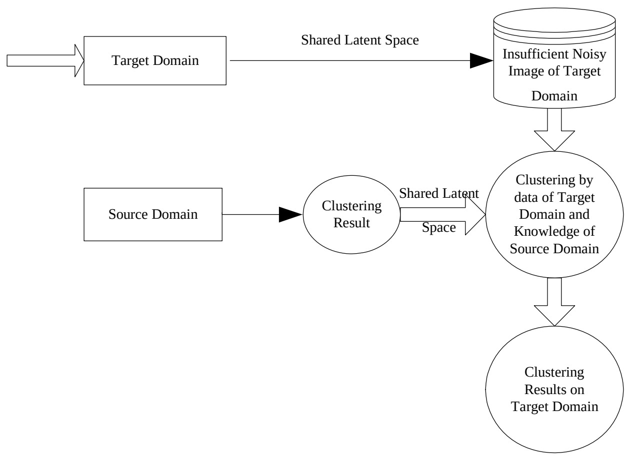

Fig.2 The schematic diagram of LSS-FTC-NTR

3.1 Negative-transfer-resistant mechanism

To resist the negative transfer, our objective is to discard bad cluster centers in the source domain and select helpful cluster centers that can help the clustering task satisfactorily. Let we have a total of N^{SD} training images in the source domain $\mathbf{x}_{i,s}(\mathbf{x}_{i,s}\in\mathbf{R}^{d\times1},i=1,2,\ldots,N^{SD})$ and N^{TD} training images in the target domain $\mathbf{x}_{i,t}(\mathbf{x}_{i,t}\in\mathbf{R}^{d\times1},i=1,2,\ldots,N^{TD})$, where $N^{TD}\ll N^{SD}$. We consider there exists a shared latent space, spanned by a projection matrix $\mathbf{\Theta}\in\mathbf{R}^{r\times d}$,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

where r is the dimensions of the shared latent space. In this way, the known cluster centers $\hat{\mathbf{v}}_h^{SD}$ (h=1, 2,..., C^{SD}) in the source domain obtained by a certain clustering algorithm could be represented as $\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}$. The projection of a target domain sample $\mathbf{x}_{i,t}$ could be represented as $\mathbf{\Theta}\mathbf{x}_{i,t}$. Suppose $\tilde{\mathbf{v}}_j^{TD}$ (j=1, 2,..., C^{TD}) is the unsolved cluster centers in the target domain in the shared latent space. We consider the following optimization term for resisting negative transfer:

$$\min_{\tilde{\mathbf{V}}} J(\tilde{\mathbf{V}}) = \sum_{j=1}^{C^{TD}}\sum_{h=1}^{C^{SD}} \left\| \tilde{\mathbf{v}}_j^{TD} - S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD} \right\|^2 , \tag{4}$$

where the parameter $S_{jh} \in \mathbf{S}^{C^{TD}\times C^{SD}}$ is called weight of transfer knowledge, and its value is in the range [0, 1]. S_{jh} denotes the matching degree between the j-th cluster center of the target domain and the h-th cluster center of the source domain. To find useful transfer knowledge from cluster centers in the source domain, it is needed to devise a strategy to set the values of S_{jh} to high values for positive transfer and low values for negative transfer. In Eq.(4), we set S_{jh} as follows:

$$S_{jh} = 1/\sum_{h'=1}^{C^{SD}} \frac{\left\| \tilde{\mathbf{v}}_j^{TD} - \mathbf{\Theta}\hat{\mathbf{v}}_h^{SD} \right\|^2}{\left\| \tilde{\mathbf{v}}_j^{TD} - \mathbf{\Theta}\hat{\mathbf{v}}_{h'}^{SD} \right\|^2} , \tag{5}$$

When S_{jh} tends to 1, $\tilde{\mathbf{v}}_j^{TD}$ exactly matches $\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}$. In this case, the clustering results will be coherent and these two centers clusters of different domains are much closer to each other. When S_{jh} tends to 0, $\tilde{\mathbf{v}}_j^{TD}$ does not match $\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}$. In this case, we make $\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}$ participate in transfer learning as little as possible. That is to say, the transfer performance will be at least no worse than performing the target clustering without transfer. In other words, if the source clustering transfer ability of $\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}$ decreases cluster performance of target domain data, it means the partial source clustering may be not related or the relationship is not sufficiently leveraged, then the negative transfer has occurred. By adjusting the value of $\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}$, Eq. (4) can help the transfer method make positive transfer when two domains are appropriately matched and resist negative transfer when two domains are not matched. At one extreme, S_{jh} is set to be 1, the transferred knowledge from the source domain are completely helpful, such that the cluster centers in the source and target domains are coincided with each other, and all transferred knowledge are completely adopted. At the other extreme, when S_{jh} tends to 0, it means the transferred knowledge from the source domain are unhelpful. The clustering on the target domain will disregard the transferred knowledge. But in most cases part transferred knowledge are selectively keep and the other parts are disregard.

3.2 The Proposed LSS-FTC-NTR

To find a proper projection matrix $\mathbf{\Theta}$, we think the difference between source and target domains in the shared latent space should be as small as possible, such that the relationship between domains is strengthened, and the transfer knowledge of data in the source domain will be more helpful to complete medical segmentation in the target domain. Based on the definition of MMD, the difference between two domains in the shared latent space can be computed as follows

$$\begin{aligned} d(P_{source}, P_{target}) &= \left\| \frac{1}{N^{TD}}\sum_{i=1}^{N^{TD}} \mathbf{\Theta}\mathbf{x}_{i,t} - \frac{1}{N^{SD}}\sum_{i=1}^{N^{SD}} \mathbf{\Theta}\mathbf{x}_{i,s} \right\|^2 \\ &= \frac{1}{(N^{TD})^2}\sum_{i=1}^{N^{TD}}\sum_{j=1}^{N^{TD}} \mathbf{\Theta}\mathbf{x}_{i,t}\mathbf{x}_{j,t}^T\mathbf{\Theta}^T + \frac{1}{(N^{SD})^2}\sum_{i=1}^{N^{SD}}\sum_{j=1}^{N^{SD}} \mathbf{\Theta}\mathbf{x}_{i,s}\mathbf{x}_{j,s}^T\mathbf{\Theta}^T \\ &\quad - \frac{2}{N^{TD}N^{SD}}\sum_{i=1}^{N^{TD}}\sum_{j=1}^{N^{SD}} \mathbf{\Theta}\mathbf{x}_{i,t}\mathbf{x}_{j,s}^T\mathbf{\Theta}^T \end{aligned} \tag{6}$$

The distribution difference between source and target domains is simply the distance between the two mean in the shared latent space. Let

$$\begin{aligned} \mathbf{\Omega} &= \frac{1}{(N^{TD})^2}\sum_{i=1}^{N^{TD}}\sum_{j=1}^{N^{TD}} \mathbf{x}_{i,t}\mathbf{x}_{j,t}^T + \frac{1}{(N^{SD})^2}\sum_{i=1}^{N^{SD}}\sum_{j=1}^{N^{SD}} \mathbf{x}_{i,s}\mathbf{x}_{j,s}^T \\ &\quad - \frac{2}{N^{TD}N^{SD}}\sum_{i=1}^{N^{TD}}\sum_{j=1}^{N^{SD}} \mathbf{x}_{i,t}\mathbf{x}_{j,s}^T \end{aligned} \tag{7}$$

The optimization of $d(P_{source}, P_{target})$ can be simplified as

$$\begin{aligned} &\min_{\mathbf{\Theta}} d(P_{source}, P_{target}) = \min_{\mathbf{\Theta}} \mathbf{\Theta}\mathbf{\Omega}\mathbf{\Theta}^T , \\ &s.t. \;\; \mathbf{\Theta}\mathbf{\Theta}^T = \mathbf{I}_{r\times r} . \end{aligned} \tag{8}$$

where $\mathbf{I}_{r\times r}$ is a $r\times r$ identity matrix, such that projection matrix $\mathbf{H}$ is orthogonal.

Coming to the transfer learning tasks, we incorporate the Eqs.(4) and (8) into the FCM framework. We obtain the objective function of LSS-FTC-NTR as follows:

$$\begin{aligned} &J = \sum_{i=1}^{N^{TD}}\sum_{j=1}^{C^{TD}} \left(\mu_{ij}^{TD}\right)^m \left(\left\| \mathbf{\Theta}\mathbf{x}_{i,t} - \tilde{\mathbf{v}}_j^{TD} \right\|^2 \right) + \lambda_1 \sum_{j=1}^{C^{TD}}\sum_{h=1}^{C^{SD}} \left\| \tilde{\mathbf{v}}_j^{TD} - S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD} \right\|^2 + \lambda_2 \mathbf{\Theta}\;\mathbf{\Omega}\mathbf{\Theta}^T \\ &s.t. \;\; \mathbf{\Theta}\mathbf{\Theta}^T = \mathbf{I}_{r\times r} , \; \sum_{j=1}^{C^{TD}} \mu_{ij}^{TD} = 1 , \end{aligned} \tag{9}$$

where parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ are the coefficients of transfer optimization term and MMD term, respectively. In LSS-FTC-NTR, the parameter λ_1 is used to control the influence of transfer optimization term $\sum_{j=1}^{C^{TD}}\sum_{h=1}^{C^{SD}} \left\| \tilde{\mathbf{v}}_j^{TD} - S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD} \right\|^2$ to the entire objective function. The larger the λ_1 value, the greater the contribution of the transfer term will be. In this case, the unsolved cluster centers $\tilde{\mathbf{v}}_j^{TD}$ in the target domain should be close to $S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}$ in the shared latent space. Conversely, when λ_1 tends to 0, the contribution of the transfer term is weakened, the difference between the unsolved cluster centers and known cluster centers in two different domains can be relaxed.

3.3 Optimization of LSS-FTC-NTR

The solution of objective function in Eq.(9) relates to the matrixes $\mathbf{\Theta}$, $\mathbf{U}$ and $\tilde{\mathbf{V}}$. In the following, we optimize them one by one using the iteratively optimization strategy. In terms of the Lagrange optimization, the minimization of J in Eq.(9) by introducing the Lagrangian multiplier α in Eq. (9) can be converted to the following unconstrained minimization problem:

$$L=\sum_{i=1}^{N^{TD}}\sum_{j=1}^{C^{TD}}\left(\mu_{ij}^{TD}\right)^m\left(\left\|\mathbf{\Theta}\mathbf{x}_{i,t}-\tilde{\mathbf{v}}_j^{TD}\right\|^2\right)+\lambda_1\sum_{j=1}^{C^{TD}}\sum_{h=1}^{C^{SD}}\left\|\tilde{\mathbf{v}}_j^{TD}-S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}\right\|^2 +\lambda_2\mathbf{\Theta}\mathbf{\Omega}\mathbf{\Theta}^T+\sum_{i=1}^{N^{TD}}\alpha_i(1-\sum_{j=1}^{C^{TD}}\mu_{ij}^{TD}). \quad (10)$$

In the first step, we fix parameters $\mathbf{\Theta}$ and $\mathbf{U}$, and only consider $\tilde{\mathbf{V}}$. To minimize this objective on parameter $\tilde{\mathbf{V}}$, we set the derivative with regard to $\tilde{\mathbf{V}}$ to zero:

$$\frac{\partial L}{\partial\tilde{\mathbf{v}}_j^{TD}}=-\sum_{i=1}^{N^{TD}}\left(\mu_{ij}^{TD}\right)^m(\mathbf{\Theta}\mathbf{x}_{i,t}-\tilde{\mathbf{v}}_j^{TD})+\lambda_1\sum_{h=1}^{C^{SD}}\left(\tilde{\mathbf{v}}_j^{TD}-S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}\right)=0$$
$$\Leftrightarrow\tilde{\mathbf{v}}_j^{TD}\left(\sum_{i=1}^{N^{TD}}\left(\mu_{ij}^{TD}\right)^m+\lambda_1C^{SD}\right)=\sum_{j=1}^{N^{TD}}\left(\mu_{ij}^{TD}\right)^m\mathbf{\Theta}\mathbf{x}_{i,t}+\lambda_1\sum_{h=1}^{C^{SD}}S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}. \quad (11)$$

We can get $\tilde{\mathbf{V}}$ in a closed form as follows

$$\tilde{\mathbf{v}}_j^{TD}=\left(\sum_{j=1}^{N^{TD}}\left(\mu_{ij}^{TD}\right)^m\mathbf{\Theta}\mathbf{x}_{i,t}+\lambda_1\sum_{h=1}^{C^{SD}}S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}\right)/\left(\sum_{i=1}^{N^{TD}}\left(\mu_{ij}^{TD}\right)^m+\lambda_1C^{SD}\right). \quad (12)$$

Likewise, in the next step, we fix parameters $\mathbf{\Theta}$ and $\tilde{\mathbf{V}}$, and only consider $\mathbf{U}$. The minimization problem of Eq.(10) with respect to $\mathbf{U}$ can be equivalent to the following problem,

$$\frac{\partial L}{\partial\mu_{ij}^{TD}}=m(\mu_{ij}^{TD})^{m-1}\left\|\mathbf{\Theta}\mathbf{x}_{i,t}-\tilde{\mathbf{v}}_j^{TD}\right\|^2-\alpha_i=0$$
$$\Leftrightarrow\mu_{ij}^{TD}=\left(\alpha_i\Big/m\left\|\mathbf{\Theta}\mathbf{x}_{i,t}-\tilde{\mathbf{v}}_j^{TD}\right\|^2\right)^{\frac{1}{m-1}}. \quad (13)$$

In light of $\sum_{j=1}^{C^{TD}}\mu_{ij}^{TD}=1$, we can obtain

$$\mu_{ij}^{TD}=\left(\frac{1}{\left\|\mathbf{\Theta}\mathbf{x}_{i,t}-\tilde{\mathbf{v}}_j^{TD}\right\|^2}\right)^{\frac{1}{m-1}}\Big/\sum_{k=1}^{C^{TD}}\left(\frac{1}{\left\|\mathbf{\Theta}\mathbf{x}_{i,t}-\tilde{\mathbf{v}}_k^{TD}\right\|^2}\right)^{\frac{1}{m-1}}. \quad (14)$$

In the next step, we update matrix $\mathbf{\Theta}$ and fix parameters $\tilde{\mathbf{V}}$ and $\mathbf{U}$. Let

$$\tilde{\mathbf{U}}_1=[\tilde{\mu}_{11},\ldots,\tilde{\mu}_{i1},\ldots,\tilde{\mu}_{N^{\mathrm{TD}}1}]\in\mathbf{R}^{1\times N^{\mathrm{TD}}}, \quad (15)$$

where

$\tilde{\mathbf{U}}=[\tilde{\mathbf{U}}_1,\ldots,\tilde{\mathbf{U}}_{C^{\mathrm{TD}}}]\in\mathbf{R}^{1\times C^{\mathrm{TD}}N^{\mathrm{TD}}}$, $\hat{\mathbf{U}}=diag(\tilde{\mathbf{U}})\in\mathbf{R}^{C^{\mathrm{TD}}N^{\mathrm{TD}}\times C^{\mathrm{TD}}N^{\mathrm{TD}}}$.

Let

$$\mathbf{\omega}=[\underbrace{\mathbf{E},\ldots,\mathbf{E}}_{C^{\mathrm{TD}}}]\in\mathbf{R}^{N^{\mathrm{TD}}\times C^{\mathrm{TD}}N^{\mathrm{TD}}}, \quad (16)$$

where $\mathbf{E}\in\mathbf{R}^{N^{\mathrm{TD}}\times N^{\mathrm{TD}}}$. Let $\mathbf{Q}=[\mathbf{q}_1,\ldots,\mathbf{q}_{C^{\mathrm{TD}}}]\in\mathbf{R}^{r\times C^{\mathrm{TD}}N^{\mathrm{TD}}}$, where $\mathbf{q}_i=[\underbrace{\mathbf{v}_i^{\mathrm{TD}},\ldots,\mathbf{v}_i^{\mathrm{TD}}}_{N^{\mathrm{TD}}}]\in\mathbf{R}^{r\times N^{\mathrm{TD}}}$. Substituting Eqs.(16), (17) and (18) back to Eq.(9), the minimization problem of Eq.(10) with respect to $\mathbf{\Theta}$ can be equivalent to the following problem,

$$\ell(\mathbf{\Theta})=tr\left((\mathbf{\Theta}\mathbf{x}_t\mathbf{\omega}-\mathbf{Q})\hat{\mathbf{U}}(\mathbf{\Theta}\mathbf{x}_t\mathbf{\omega}-\mathbf{Q})^T\right) +\lambda_1tr\left((\mathbf{\Theta}\hat{\mathbf{V}}^{\mathrm{SD}}\mathbf{S}^T-\tilde{\mathbf{V}}^{\mathrm{TD}})(\mathbf{\Theta}\hat{\mathbf{V}}^{\mathrm{SD}}\mathbf{S}^T-\tilde{\mathbf{V}}^{\mathrm{TD}})^T\right)+\lambda_2tr(\mathbf{\Theta}\mathbf{\Omega}\mathbf{\Theta}^T), \quad (17)$$

where

$$\mathbf{\Omega}=\frac{1}{(N^{TD})^2}\mathbf{x}_t\left[\mathbf{1}\right]^{N^{TD}\times N^{TD}}\left(\mathbf{x}_t\right)^T+\frac{1}{(N^{SD})^2}\mathbf{x}_s\left[\mathbf{1}\right]^{N^{SD}\times N^{SD}}\left(\mathbf{x}_s\right)^T -\frac{1}{N^{TD}N^{SD}}\mathbf{x}_t\left[\mathbf{1}\right]^{N^{TD}\times N^{SD}}\left(\mathbf{x}_s\right)^T-\frac{1}{N^{TD}N^{SD}}\mathbf{x}_s\left[\mathbf{1}\right]^{N^{SD}\times N^{TD}}\left(\mathbf{x}_t\right)^T.$$

Likewise, the minimization problem of Eq.(10) with respect to $\mathbf{\Theta}$ can be equivalent to the following problem,

$$\frac{\partial\ell}{\partial\mathbf{\Theta}}=(\mathbf{\Theta}\mathbf{x}_t\mathbf{\omega}\hat{\mathbf{U}}\left(\mathbf{\omega}\right)^T\left(\mathbf{x}_t\right)^T-\mathbf{Q}\hat{\mathbf{U}}\left(\mathbf{\omega}\right)^T\left(\mathbf{x}_t\right)^T) +\lambda_1\left(\mathbf{\Theta}\hat{\mathbf{V}}^{\mathrm{SD}}\mathbf{S}^T\mathbf{S}\left(\hat{\mathbf{V}}^{\mathrm{SD}}\right)^T-\tilde{\mathbf{V}}^{\mathrm{TD}}\mathbf{S}\left(\hat{\mathbf{V}}^{\mathrm{SD}}\right)^T\right)+\lambda_2\mathbf{\Theta}\mathbf{\Omega}. \quad (18)$$

In this study, the widely gradient descent method is adopted to compute the optimal $\mathbf{\Theta}$. By setting the initial value $\mathbf{\Theta}$ as $\mathbf{\Theta}^0$, the gradient descent method successively optimal $\mathbf{\Theta}$ as follows

$$\mathbf{\Theta}^l=\mathbf{\Theta}^{l-1}-\eta\frac{\partial\ell}{\partial\mathbf{\Theta}}\Big|_{\mathbf{\Theta}=\mathbf{\Theta}^{l-1}}, \quad (19)$$

where η is the learning rate and l is the iteration number. Considering the constraint of $\mathbf{\Theta}\mathbf{\Theta}^T=\mathbf{I}$. After each updating step of $\mathbf{\Theta}^l$, let $\mathbf{\Theta}^l=\mathbf{\Theta}^{l'}\mathbf{R}$ be the QR decomposition of $\mathbf{\Theta}^l$, where $\mathbf{\Theta}^{l'}$ has orthogonal columns and $\mathbf{R}$ is an upper triangle. Then we replace $\mathbf{\Theta}^l$ with $\mathbf{\Theta}^{l'}$ for the next iteration. Eq. (19) will be iteratively solved until the convergence condition is satisfied.

Based on the above analysis, the proposed LSS-FTC-NTR model is presented in Algorithm 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Algorithm 1: LSS-FTC-NTR model	
Initialize	Set the maximum number of iterations t_{max}, the fuzzy index m, the regularization parameters λ_1 and λ_2, and the learning rate η .
Repeat:	
Exacting transfer knowledge form the source domain;	
Perform soft-partition clustering methods in the source domain, such as FCM, and obtain the cluster centers of data in the source domain;	
$t = t$+1;	
Initialize the clustering centers of data in the target domain;	
Compute the weight of transfer knowledge S_{jh} using Eq. (10);	
Fix $\mathbf{U}(t)$ and $\mathbf{\Theta}(t)$, obtain $\tilde{\mathbf{V}}^{TD}(t)$ using Eq. (12);	
Fix $\tilde{\mathbf{V}}^{TD}(t)$ and $\mathbf{\Theta}(t)$, obtain $\mathbf{U}(t)$ using Eq. (14);	
Fix $\mathbf{U}(t)$ and $\tilde{\mathbf{V}}^{TD}(t)$, obtain $\mathbf{\Theta}(t)$ using Eq. (18) and Eq.(19);	
Compute $J(t)$ using Eq. (9);	
Until $\|J(t)-J(t-1)\| \le \delta$ or $t \ge t_{max}$;	

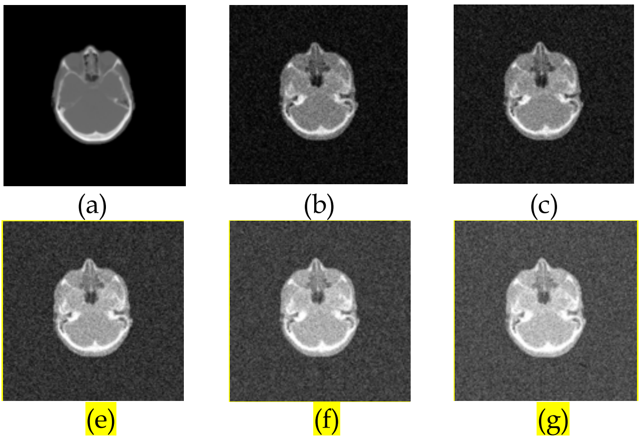

Fig. 3. The example brain CT images in source and target domains, (a) images in source domain, (b) Subject1 in target domain with 5% noise, (c) Subject1 in target domain with 10% noise, (d) Subject1 in target domain with 15% noise, (e) Subject1 in target domain with 20% noise, (f) Subject1 in target domain with 25% noise, (g) Subject1 in target domain with 30% noise

4 Experiments

4.1 Data Sets and Settings

We use ultrashort echo time (UTE) and modified Dixon brain image datasets [39, 40]. It consists of 256 brain CT image slices of 10 patients, with each image of a resolution of 256 x 256 pixels. All CT images with corresponding manual segmentation are segmented into three classes: bone, water and soft issues. These class labels are assigned by physicians or technicians. We randomly select 20 brain CT images as the original target domain data, and the rest 236 brain CT images as source domain data. We consider the application of LSS-FTC-NTR in the scenario of target images polluted by noise. To this aim, all target images were corrupted by 5%, 10%, 15%, 20%, 25% and 30% Gaussian noise. The example images in the source and target domains are shown in Fig.3. Following the training protocol established in [41], we construct a total training data set combining 236 source brain images and random 8 target brain images, while the remaining 12 target brain images are used as testing brain images. We repeat the experiment for 10 runs and record the experimental results.

To compare the segmentation performance of LSS-FTC-NTR with that of existing methods, complete image segmentation is obtained and compared with segmentations obtained by FCM [27], transfer spectral clustering (TSC) [24], and type-I knowledge-transfer-oriented *c*-means (T1-KT-FCM) [28]. As introduced in section II, FCM is the baseline algorithm of LSS-FTC-NTR. TSC performs transfer learning based on bipartite graph co-clustering, which adopts both the data manifold and sample manifold shared among different domains. T1-KT-FCM makes the cluster centers in the source domain as the transfer information to control the knowledge transfer in the test images. T1-KT-FCM incorporates this idea into FCM to achieve automatic image segmentation. To obtain the optimal parameters in all four methods, the common used grid search is conducted. Fuzzy index m in all fuzzy clusters is set within the grid $\{1.1, 1.5, 2, 2.5\}$. The K-nearest parameters in TSC are set within the grid {0, 0.005, 0.1, 0.5, 0.7, 1, 1.5, 10, 50, 100}. The λ, γ parameters in T1-KT-FCM are set within the grid $\{10e\text{-}4, 10e\text{-}3, \ldots, 10e4\}$. The parameters λ_1, λ_2 in LSS-FTC-NTR are set within the grid $\{0, 10e\text{-}4, 10e\text{-}3, \ldots, 10e6\}$, learning rate η in LSS-FTC-NTR are set within the grid $\{1e\text{-}4, 1e\text{-}3, 1e\text{-}2\}$, and the maximum number of iterations is 10e5.

In this study, the performance of image segmentation by clustering methods is evaluated in terms of two validity indicators: normalized mutual information (NMI) [42] and adjusted rand index (ARI) [43]. NMI and ARI can efficiently evaluate the agreement degree between the known clusters and the estimated data structure. Both NMI and ARI take values from 0 to 1, and larger value means better cluster performance. Experimental environment is Intel Core i3-4170 3.7GHz CPU and 12GM RAM, Windows 10, and MATLAB R2016a in this study.

4.2 Performance Comparison

The clustering performance of four methods is reported in the following. The mean and standard deviation of NMI and ARI for all compared clustering methods are displayed in Tables 1-6, respectively. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

experimental results show that three transfer learning methods are superior to FCM. The introduction of transfer knowledge from source domain has indeed promoted the cluster performance of data in the target domain. FCM is not a transfer learning cluster method which simply combines the source domain and target domain data as the training data. Due to underlying noise or outliers in the target domain, the distribution difference between source and target domains are significant different. Thus, FCM can not obtain good clustering performance in terms of NMI and ARI. Our model achieves the best performance in all datasets. TSC may be not suitable for the transfer scenario in noisy medical image segmentation, since the character of medical image are usually different in noisy scenario, while the manifold and sample manifold shared among different domains can not resist negative transfer for TSC. T1-KT-FCM exploits the transfer knowledge across domains in the original data space; however, the limited transfer knowledge can not be fully exploited in such original space. LSS-FTC-NTR has shown better performance than the other comparison methods in terms of NMI and ARI. Both the reliable knowledge obtained in the source domain and the ability of resisting negative transfer has the important influence on the segmentation performance of LSS-FTC-NTR. To better observe the behavior of all algorithms, Figs. 4-9 graphically shows the segmentation results of all comparison methods obtained on subject1 with different noise. Similar to the results in the Tables 1-12, LSS-FTC-NTR obtains the best segmentation results for distinguishing the bone, water and soft issues. The boundaries between different organizations are smooth, and obvious are relatively clearer than the other three methods.

Table 1
NMI performance of all comparison methods on 5% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0112	0.0041	0.5085	0.2249	0.5802	0.0254	**0.7146**	0.0056
Subject2	0.1515	0.1468	0.5801	0.1632	0.6359	0.0148	**0.7526**	0.0142
Subject3	0.1583	0.1040	0.5573	0.1591	0.6438	0.0290	**0.7253**	0.0059
Subject4	0.0137	0.0062	0.5149	0.1868	0.6091	0.0590	**0.7256**	0.0138
Subject5	0.2333	0.1355	0.5649	0.1840	0.6329	0.0254	**0.7432**	0.0094
Subject6	0.1823	0.1094	0.5659	0.1446	0.6505	0.0234	**0.7441**	0.0116
Subject7	0.0086	0.0046	0.5283	0.2242	0.6096	0.0212	**0.7352**	0.0110
Subject8	0.1324	0.1157	0.5664	0.1917	0.6603	0.0245	**0.7586**	0.0108
Subject9	0.1188	0.1240	0.5845	0.1638	0.6700	0.0399	**0.7693**	0.0111
Subject10	0.0185	0.0083	0.5765	0.2749	0.6512	0.0258	**0.7781**	0.0127
Subject11	0.1178	0.2362	0.5882	0.2124	0.6580	0.0179	**0.7862**	0.0068
Subject12	0.0642	0.0376	0.5719	0.1542	0.6631	0.0358	**0.7759**	0.0079

Table 2
NMI performance of all comparison methods on 10% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0155	0.0098	0.5063	0.2137	0.4748	0.0095	**0.6337**	0.0091
Subject2	0.2943	0.1666	0.5797	0.1689	0.5792	0.0194	**0.7230**	0.0045
Subject3	0.2182	0.0813	0.5502	0.1716	0.5224	0.0366	**0.6908**	0.0117
Subject4	0.0196	0.0163	0.5061	0.1966	0.4706	0.0110	**0.6475**	0.0127
Subject5	0.1959	0.1420	0.5635	0.1591	0.5273	0.0205	**0.7056**	0.0122
Subject6	0.2868	0.0777	0.5683	0.1421	0.5310	0.0162	**0.6894**	0.0177
Subject7	0.0076	0.0038	0.5336	0.2612	0.4696	0.0106	**0.6710**	0.0162
Subject8	0.2006	0.1377	0.5680	0.1448	0.5357	0.0217	**0.7077**	0.0158
Subject9	0.2052	0.1502	0.5923	0.1901	0.5409	0.0165	**0.7274**	0.0104
Subject10	0.0153	0.0136	0.5755	0.2609	0.5103	0.0186	**0.7244**	0.0046
Subject11	0.0814	0.1664	0.5846	0.2263	0.5231	0.0215	**0.7203**	0.0122
Subject12	0.1534	0.1036	0.5694	0.1864	0.5409	0.0062	**0.7144**	0.0149

Table 3
NMI performance of all comparison methods on 15% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0130	0.0106	0.5053	0.2261	0.4470	0.0181	**0.5684**	0.0082
Subject2	0.2346	0.1282	0.5815	0.1536	0.5631	0.0172	**0.6452**	0.0118
Subject3	0.2155	0.0679	0.5541	0.2003	0.4881	0.0175	**0.5978**	0.0024
Subject4	0.1789	0.2392	0.5076	0.2182	0.4509	0.0171	**0.5849**	0.0157
Subject5	0.2061	0.0862	0.5621	0.1394	0.5341	0.0121	**0.6220**	0.0030
Subject6	0.2592	0.1368	0.5691	0.1526	0.5079	0.0163	**0.6261**	0.0096
Subject7	0.0963	0.1848	0.5274	0.1991	0.4664	0.0168	**0.5939**	0.0130
Subject8	0.1826	0.1223	0.5674	0.1458	0.5087	0.0223	**0.6198**	0.0101
Subject9	0.2245	0.1570	0.5869	0.1486	0.5118	0.0120	**0.6419**	0.0157
Subject10	0.0735	0.1450	0.5697	0.2271	0.4975	0.0197	**0.6098**	0.0169
Subject11	0.0937	0.1846	0.5826	0.2770	0.5267	0.0242	**0.6107**	0.0075
Subject12	0.2454	0.1816	0.5720	0.2135	0.5328	0.0129	**0.6537**	0.0117

Table 4
NMI performance of all comparison methods on 20% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0246	0.0289	0.5095	0.1908	0.4782	0.0069	**0.6000**	0.0055
Subject2	0.3092	0.1299	0.5796	0.1631	0.5596	0.0082	**0.6348**	0.0128
Subject3	0.2597	0.0992	0.5556	0.1588	0.5029	0.0108	**0.5926**	0.0051
Subject4	0.0102	0.0061	0.5136	0.2026	0.4778	0.0066	**0.5254**	0.0133
Subject5	0.3071	0.0852	0.5654	0.1471	0.5333	0.0119	**0.6237**	0.0084
Subject6	0.2998	0.0697	0.5662	0.1376	0.5247	0.0176	**0.5875**	0.0099
Subject7	0.0994	0.1742	0.5298	0.2377	0.4804	0.0088	**0.5719**	0.0063
Subject8	0.2557	0.1427	0.5683	0.1438	0.5426	0.0160	**0.6023**	0.0114
Subject9	0.3098	0.0342	0.5852	0.1508	0.5366	0.0089	**0.6433**	0.0062
Subject10	0.0966	0.2007	0.5788	0.2565	0.5309	0.0066	**0.5991**	0.0080
Subject11	0.1196	0.2075	0.5818	0.2101	0.5525	0.0080	**0.6516**	0.0129
Subject12	0.1530	0.1197	0.5707	0.1767	0.5466	0.0167	**0.5869**	0.0122

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Table 5
NMI performance of all comparison methods on 25% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0107	0.0058	0.5078	0.2154	0.4969	0.0043	**0.5469**	0.0104
Subject2	0.1708	0.1001	0.5792	0.1741	0.5842	0.0076	**0.6572**	0.0182
Subject3	0.2460	0.0981	0.5503	0.1685	0.5232	0.0145	**0.5940**	0.0093
Subject4	0.0121	0.0081	0.5177	0.2390	0.5045	0.0112	**0.5717**	0.0197
Subject5	0.1503	0.1169	0.5621	0.1599	0.5580	0.0122	**0.6090**	0.0083
Subject6	0.2199	0.0660	0.5662	0.1533	0.5434	0.0311	**0.5935**	0.0142
Subject7	0.0284	0.0285	0.5331	0.2528	0.5156	0.0137	**0.5677**	0.0095
Subject8	0.1320	0.1375	0.5705	0.1577	0.5547	0.0156	**0.6272**	0.0136
Subject9	0.2594	0.1054	0.5859	0.1542	0.5682	0.0124	**0.6437**	0.0127
Subject10	0.1980	0.0293	0.5763	0.2203	0.5488	0.0073	**0.6416**	0.0157
Subject11	0.2156	0.2697	0.5830	0.2349	0.5745	0.0184	**0.6732**	0.0117
Subject12	0.3935	0.0908	0.5724	0.1827	0.5617	0.0122	**0.6386**	0.0078

Table 6
NMI performance of all comparison methods on 30% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0120	0.0093	0.5086	0.2028	0.5569	0.0124	**0.6567**	0.0084
Subject2	0.2553	0.1770	0.5833	0.1497	0.6333	0.0113	**0.7439**	0.0109
Subject3	0.2010	0.0804	0.5526	0.1537	0.5872	0.0206	**0.6602**	0.0110
Subject4	0.0891	0.1752	0.5143	0.2144	0.5664	0.0050	**0.6720**	0.0114
Subject5	0.1820	0.1376	0.5663	0.1463	0.6172	0.0193	**0.7080**	0.0028
Subject6	0.2316	0.1094	0.5695	0.1928	0.6062	0.0120	**0.6872**	0.0086
Subject7	0.1080	0.1686	0.5316	0.2091	0.5769	0.0172	**0.6966**	0.0054
Subject8	0.1755	0.1045	0.5696	0.1777	0.6231	0.0097	**0.7037**	0.0103
Subject9	0.2584	0.0893	0.5913	0.1901	0.6295	0.0216	**0.7356**	0.0200
Subject10	0.0113	0.0135	0.5751	0.2901	0.6247	0.0198	**0.7441**	0.0115
Subject11	0.2121	0.2838	0.5827	0.2652	0.6343	0.0212	**0.7491**	0.0042
Subject12	0.1547	0.1170	0.5702	0.1941	0.6226	0.0045	**0.7256**	0.0095

Table 7
ARI performance of all comparison methods on 5% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0021	0.0043	0.3399	0.1756	0.7579	0.0301	**0.8880**	0.0028
Subject2	0.0854	0.0814	0.3931	0.1561	0.8032	0.0160	**0.9111**	0.0109
Subject3	0.0861	0.0521	0.4144	0.1478	0.8072	0.0318	**0.8839**	0.0031
Subject4	0.0013	0.0041	0.3466	0.1701	0.7730	0.0629	**0.8913**	0.0107
Subject5	0.1186	0.0877	0.3884	0.1761	0.8085	0.0316	**0.9080**	0.0078
Subject6	0.0839	0.0560	0.4153	0.1393	0.8069	0.0264	**0.8950**	0.0108
Subject7	0.0012	0.0026	0.3593	0.1594	0.7728	0.0243	**0.8922**	0.0119
Subject8	0.0651	0.0732	0.3902	0.1753	0.8293	0.0276	**0.9169**	0.0095
Subject9	0.0658	0.0682	0.4153	0.1524	0.8305	0.0378	**0.9187**	0.0102
Subject10	0.0071	0.0055	0.3806	0.2090	0.8036	0.0304	**0.9242**	0.0106
Subject11	0.0771	0.1614	0.3858	0.1903	0.8089	0.0215	**0.9299**	0.0053
Subject12	0.0094	0.0109	0.3744	0.1365	0.8131	0.0424	**0.9269**	0.0081

Table 8
ARI performance of all comparison methods on 10% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0028	0.0113	0.3403	0.1791	0.5742	0.0284	**0.8155**	0.0056
Subject2	0.1798	0.1182	0.3955	0.1613	0.5515	0.0292	**0.8812**	0.0056
Subject3	0.1158	0.0542	0.4121	0.1696	0.6163	0.0635	**0.8515**	0.0148
Subject4	0.0046	0.0063	0.3433	0.1729	0.5650	0.0237	**0.8281**	0.0134
Subject5	0.1051	0.0953	0.3848	0.1535	0.6339	0.0317	**0.8777**	0.0093
Subject6	0.1523	0.0812	0.4168	0.1373	0.6359	0.0297	**0.8467**	0.0136
Subject7	0.0008	0.0012	0.3607	0.1920	0.5573	0.0211	**0.8482**	0.0125
Subject8	0.1140	0.0988	0.3915	0.1311	0.6423	0.0301	**0.8522**	0.0149
Subject9	0.1202	0.1143	0.4210	0.1781	0.6382	0.0300	**0.8839**	0.0101
Subject10	0.0011	0.0089	0.3800	0.1963	0.5939	0.0328	**0.8849**	0.0037
Subject11	0.0493	0.1062	0.3848	0.1608	0.6045	0.0401	**0.8742**	0.0118
Subject12	0.0484	0.0401	0.3737	0.1681	0.6302	0.0112	**0.8738**	0.0125

Table 9
ARI performance of all comparison methods on 15% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0009	0.0031	0.3389	0.1753	0.3607	0.0262	**0.6833**	0.0089
Subject2	0.1206	0.1058	0.3935	0.1414	0.4404	0.0126	**0.6679**	0.0107
Subject3	0.1083	0.0540	0.4127	0.1807	0.4323	0.0238	**0.5997**	0.0032
Subject4	0.1239	0.1680	0.3439	0.1746	0.3620	0.0344	**0.7034**	0.0137
Subject5	0.0847	0.0405	0.3859	0.1306	0.4408	0.0298	**0.7278**	0.0046
Subject6	0.1421	0.1255	0.4182	0.1430	0.4259	0.0204	**0.6636**	0.0077
Subject7	0.0639	0.1304	0.3588	0.1519	0.3662	0.0287	**0.5863**	0.0161
Subject8	0.0729	0.0824	0.3902	0.1349	0.4120	0.0355	**0.5671**	0.0136
Subject9	0.1302	0.0900	0.4174	0.1444	0.4186	0.0102	**0.6287**	0.0148
Subject10	0.0398	0.0848	0.3778	0.2008	0.4098	0.0241	**0.5256**	0.0122
Subject11	0.0593	0.1239	0.3843	0.2049	0.4230	0.0250	**0.6560**	0.0087
Subject12	0.1478	0.1185	0.3750	0.1692	0.4119	0.0303	**0.6518**	0.0128

Table 10
ARI performance of all comparison methods on 20% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0019	0.0080	0.3400	0.1486	0.3241	0.0059	**0.6624**	0.0098
Subject2	0.1754	0.1208	0.3933	0.1522	0.3859	0.0059	**0.6005**	0.0116
Subject3	0.1443	0.0831	0.4147	0.1533	0.3742	0.0102	**0.5570**	0.0066
Subject4	0.0015	0.0060	0.3464	0.1713	0.3270	0.0090	**0.5008**	0.0139
Subject5	0.1679	0.0931	0.3883	0.1370	0.3730	0.0056	**0.6152**	0.0068
Subject6	0.1701	0.0682	0.4176	0.1304	0.3838	0.0091	**0.5775**	0.0079
Subject7	0.0625	0.1226	0.3602	0.1912	0.3309	0.0064	**0.4983**	0.0083
Subject8	0.1520	0.1066	0.3913	0.1339	0.3809	0.0088	**0.5055**	0.0102
Subject9	0.1619	0.0404	0.4164	0.1424	0.3871	0.0056	**0.5725**	0.0053
Subject10	0.0620	0.1394	0.3812	0.1927	0.3571	0.0027	**0.5494**	0.0065
Subject11	0.0713	0.1457	0.3838	0.1649	0.3695	0.0065	**0.5819**	0.0122
Subject12	0.0657	0.0780	0.3746	0.1617	0.3641	0.0048	**0.4376**	0.0106

Table 11
ARI performance of all comparison methods on 25% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0023	0.0020	0.3409	0.1786	0.3419	0.0056	**0.5128**	0.0139
Subject2	0.0687	0.0687	0.3948	0.1626	0.4354	0.0132	**0.6609**	0.0156
Subject3	0.1447	0.0664	0.4126	0.1659	0.4043	0.0113	**0.5219**	0.0102
Subject4	0.0125	0.0047	0.3484	0.1856	0.3738	0.0138	**0.5349**	0.0204
Subject5	0.0701	0.0553	0.3862	0.1545	0.4182	0.0235	**0.6445**	0.0077
Subject6	0.1147	0.0392	0.4186	0.1502	0.4153	0.0305	**0.5271**	0.0151
Subject7	0.0076	0.0109	0.3611	0.1885	0.3654	0.0088	**0.5132**	0.0116
Subject8	0.0656	0.0666	0.3914	0.1499	0.4143	0.0176	**0.5902**	0.0148
Subject9	0.1248	0.0841	0.4173	0.1462	0.4233	0.0152	**0.6459**	0.0121
Subject10	0.1157	0.0035	0.3804	0.1958	0.3796	0.0046	**0.5953**	0.0165
Subject11	0.1382	0.1889	0.3842	0.1686	0.4277	0.0200	**0.6537**	0.0123
Subject12	0.2421	0.0869	0.3739	0.1749	0.4058	0.0177	**0.6031**	0.0091

Table 12
ARI performance of all comparison methods on 30% noisy CT image datasets

Dataset	FCM		TSC		T1-KT-FCM		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0008	0.0052	0.3401	0.1505	0.5432	0.0294	**0.7743**	0.0062
Subject2	0.1663	0.1180	0.3941	0.1505	0.5887	0.0309	**0.8122**	0.0136
Subject3	0.0947	0.0389	0.4132	0.1584	0.5700	0.0397	**0.7414**	0.0125
Subject4	0.0601	0.1242	0.3471	0.1760	0.5526	0.0094	**0.7876**	0.0146
Subject5	0.0867	0.0967	0.3886	0.1351	0.5858	0.0488	**0.7966**	0.0044
Subject6	0.1193	0.0774	0.4173	0.1744	0.5833	0.0131	**0.7834**	0.0053
Subject7	0.0675	0.1173	0.3611	0.1547	0.5637	0.0408	**0.7874**	0.0079
Subject8	0.0684	0.0490	0.3911	0.1728	0.6018	0.0277	**0.7363**	0.0144
Subject9	0.1227	0.0787	0.4184	0.1823	0.6060	0.0214	**0.8233**	0.0176
Subject10	0.0047	0.0076	0.3800	0.2041	0.5840	0.0421	**0.8140**	0.0134
Subject11	0.1420	0.1947	0.3839	0.2030	0.6010	0.0318	**0.8226**	0.0052
Subject12	0.0588	0.0640	0.3745	0.1792	0.5793	0.0106	**0.7996**	0.0093

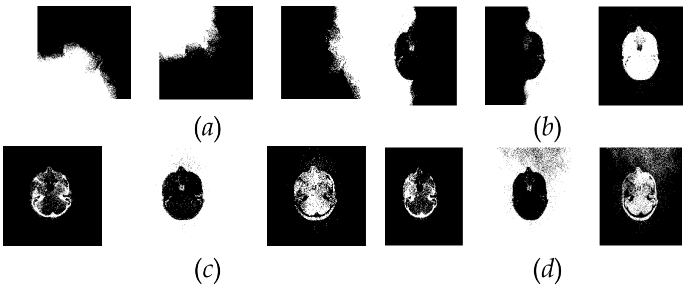

Fig.4 Clustering segmentations on subject1+5% noise, (a)FCM, (b)TSC, (c)T1-KT-FCM, (d)LSS-FTC-NTR

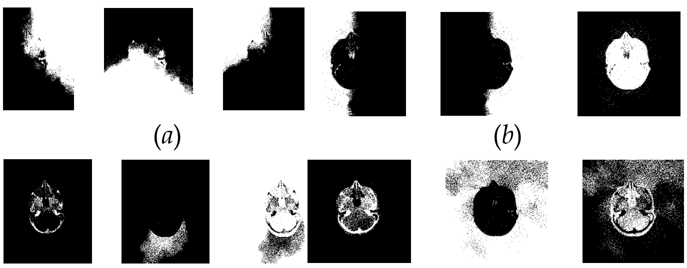

Fig.5 Clustering segmentations on subject1+10% noise, (a)FCM, (b)TSC, (c)T1-KT-FCM, (d)LSS-FTC-NTR

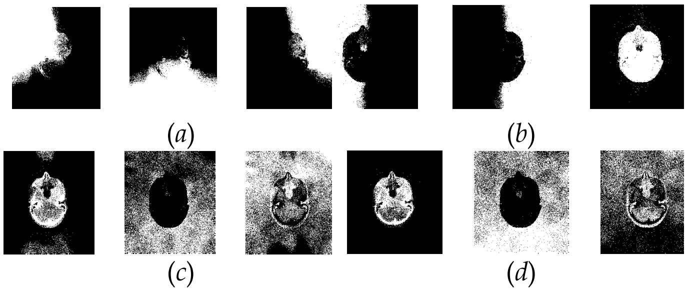

Fig.6 Clustering segmentations on subject1+15% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

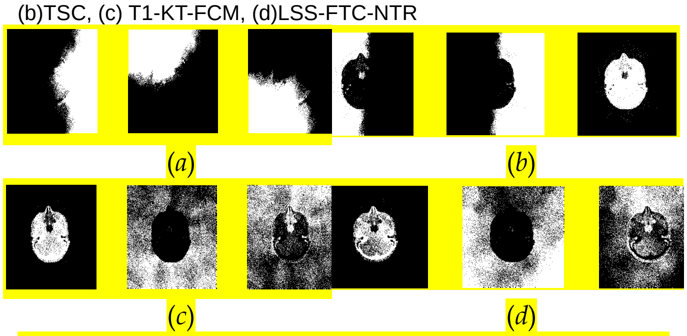

Fig.7 Clustering segmentations on subject1+20% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

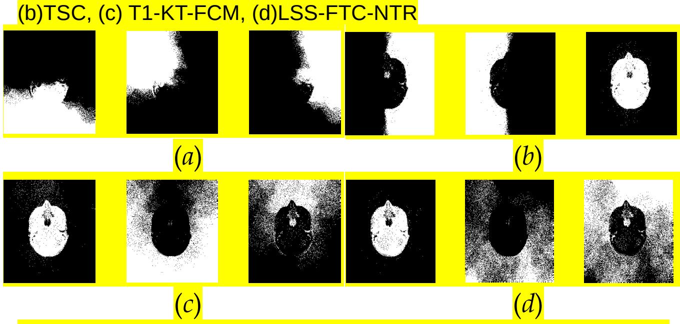

Fig.8 Clustering segmentations on subject1+25% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

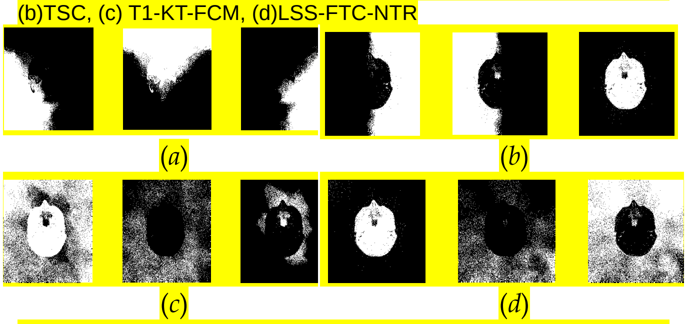

Fig.9 Clustering segmentations on subject1+30% noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)LSS-FTC-NTR

4.3 Flexibility Evaluation of LSS-FTC-NTR

To validate the effect of two regularization terms on the performance of LSS-FTC-NTR, we present two comparison methods LSS-FTC-NTR (λ_1=0) and LSS-FTC-NTR (λ_2=0), obtained with the parameter λ_1=0 and λ_2=0 in LSS-FTC-NTR, respectively. We compare them with FCM and LSS-FTC-NTR on Subjects 1-8 and show their mean and standard deviation of NMI and ARI in Tables 13-14, respectively. The experimental results show that the performances of both LSS-FTC-NTR (λ_1=0) and LSS-FTC-NTR (λ_2=0) are better than baseline method FCM. The regularization term in LSS-FTC-NTR (λ_2=0) is $\sum_{j=1}^{C^{TD}}\sum_{h=1}^{C^{SD}}\left\|\tilde{\mathbf{v}}_j^{TD}-S_{jh}\mathbf{\Theta}\hat{\mathbf{v}}_h^{SD}\right\|^2$, which can effectively resist negative transform by using the transfer optimization strategy and improve the segmentation performance in noisy scenario. The regularization term in LSS-FTC-NTR (λ_1=0) is $\mathbf{\Theta}^T\mathbf{\Omega}\mathbf{\Theta}$, which finds a shared latent space for data cross domains, so that the projection data distributions of the source and target domains are close to each other. LSS-FTC-NTR has the advantages of both LSS-FTC-NTR (λ_1=0) and LSS-FTC-NTR (λ_2=0). It can exploit more transfer knowledge; meanwhile, and achieve a good balance between making use of positive transfer and resisting negative transfer.

Table 13
NMI performance of all comparison methods on 5% noisy CT image datasets

Dataset	FCM		LSS-FTC-NTR (λ_1=0)		LSS-FTC-NTR (λ_2=0)		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0112	0.0041	0.1312	0.0087	0.7009	0.0101	**0.7146**	0.0056
Subject2	0.1515	0.1468	0.4187	0.0121	0.7483	0.0135	**0.7526**	0.0142
Subject3	0.1583	0.1040	0.3852	0.0099	0.7011	0.0074	**0.7253**	0.0059
Subject4	0.0137	0.0062	0.1474	0.0137	0.7088	0.0062	**0.7256**	0.0138
Subject5	0.2333	0.1355	0.3566	0.0173	0.7304	0.0121	**0.7432**	0.0094
Subject6	0.1823	0.1094	0.4366	0.0075	0.7231	0.0071	**0.7441**	0.0116
Subject7	0.0086	0.0046	0.2283	0.0108	0.7184	0.0069	**0.7352**	0.0110
Subject8	0.1324	0.1157	0.4831	0.0078	0.7389	0.0082	**0.7586**	0.0108

Table 14
ARI performance of all comparison methods on 5% noisy CT image datasets

Dataset	FCM		LSS-FTC-NTR (λ_1=0)		LSS-FTC-NTR (λ_2=0)		LSS-FTC-NTR	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0021	0.0043	0.4671	0.0069	0.8621	0.0042	**0.8880**	0.0028
Subject2	0.0854	0.0814	0.6893	0.0114	0.9017	0.0089	**0.9111**	0.0109
Subject3	0.0861	0.0521	0.7022	0.0093	0.8702	0.0047	**0.8839**	0.0031
Subject4	0.0013	0.0041	0.4705	0.0078	0.8856	0.0122	**0.8913**	0.0107
Subject5	0.1186	0.0877	0.6244	0.0086	0.8901	0.0099	**0.9080**	0.0078
Subject6	0.0839	0.0560	0.5921	0.0077	0.8901	0.0058	**0.8950**	0.0108
Subject7	0.0012	0.0026	0.5156	0.0143	0.8815	0.0103	**0.8922**	0.0119
Subject8	0.0651	0.0732	0.5702	0.0084	0.8977	0.0127	**0.9169**	0.0095

Next, we discuss the influence of the number of samples in the source domain on the performance of LSS-FTC-NTR. We randomly select 10%, 30%, 50%, 70%, 90% and 100% proportion of training samples in the source domain as the source training dataset. To make the results fair, we repeat the above sampling 10 times for each sample size. NMI and ARI performances of LSS-FTC-NTR on Subject1 and Subject2 with 5% noisy are shown in Fig.10 and Fig.11, respectively. The experimental results show that the values of NMI and ARI increase with the increase of the number of samples in the source domain. The reason is that LSS-FTC-NTR can not mine enough transfer knowledge from source domain when training samples in the source domain are too few. On the other hand, exploiting clear and concise transfer knowledge need a certain amount of high quality samples in the source domain. Thus, it can be inferred that the more samples in the source domain, the more helpful the knowledge obtained in the source domain and the more efficient the LSS-FTC-NTR will be in the target domain.

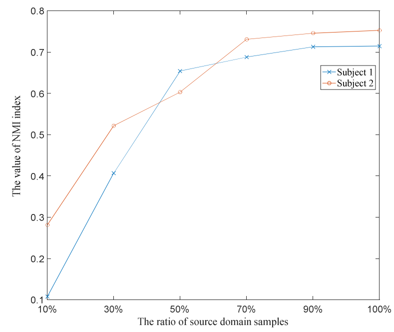

Fig.10 NMI performance of LSS-FTC-NTR with different proportion of samples in the source domain on 5% noisy Subject1 and Subject2

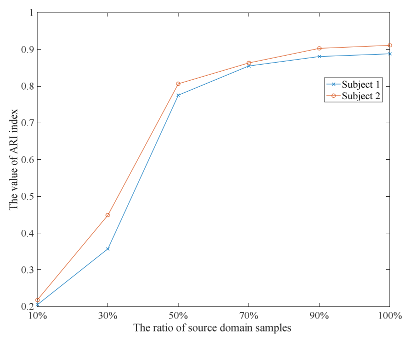

Fig.11 ARI performance of LSS-FTC-NTR with different proportion of samples in the source domain on 5% noisy Subject1 and Subject2

4.3 Parameter Sensitive

In the experiments, the parameters λ_1 and λ_2 are determined in a given search grid. In the following, we discuss the performance of LSS-FTC-NTR using different parameters. Tables 15-16 show the means of NMI and ARI on the subject using different λ_1 and λ_2, while fixing the parameter m=2.

1) LSS-FTC-NTR is sensitive to parameters λ_1 and λ_2. Different λ_1 and λ_2 lend to different cluster performance of LSS-FTC-NTR in terms of NMI and ARI. It can be found that in most situations when the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2963873, IEEE/ACM Transactions on Computational Biology and Bioinformatics

value of NMI is better, the value of ARI is also better. Thus, it is feasible to use NMI and ARI as performance criterions to determine the suitable parameters.

2) Fixed the value of *m*, LSS-FTC-NTR obtains the worst NMI and ARI when $\lambda_1 = 0$ and $\lambda_2 = 0$. The clustering performance of LSS-FTC-NTR is improved when λ_1 and λ_2 are not equal to 0. Since when $\lambda_1 = 0$ and $\lambda_2 = 0$ LSS-FTC-NTR is degenerated to the classical FCM clustering.

3) We can find that when the value of λ_1 is large, LSS-FTC-NTR obtains the satisfactory performance in terms of NMI and ARI. This further demonstrates that the proposed negative-transfer-resistance mechanism has played an effective role. Thus, in the subsequent experiments, we can reduce the search grid of λ_1 in the range $\{10e1, 10e1, \ldots, 10e6\}$. We can't find the rule to select parameter λ_2. We think it is reasonable to select optimal λ_2 within the search grid. The range $\lambda_2 \in \{10e\text{-}4, 10e\text{-}3, \ldots, 10e6\}$ is appropriate.

Table 15

Means of NMI by LSS-FTC-NTR on the subject1+5% noise using different λ_1 and λ_2, while fixing *m*=2

λ_2 \ λ_1	0	10e-4	10e-3	10e-2	10e-1	1	10e1	10e2	10e3	10e4	10e5	10e6
0	0.4801	0.5006	0.5181	0.5501	0.6011	0.6250	0.6091	0.6375	0.6788	0.6397	0.6427	0.6378
10e-4	0.4915	0.5326	0.5592	0.5662	0.6109	0.6161	0.6469	0.6499	0.6468	0.6493	0.6431	0.6425
10e-3	0.5094	0.5436	0.5572	0.5866	0.7007	0.6540	0.6413	0.6456	0.6465	0.6457	0.6449	0.6438
10e-2	0.5054	0.5605	0.5449	0.6980	**0.7159**	0.7006	0.6971	0.7075	0.6905	0.6766	0.7017	0.7106
10e-1	0.5036	0.5025	0.5036	0.5017	0.5070	0.5140	0.5138	0.5216	0.5184	0.5140	0.5147	0.5195
1	0.3540	0.3162	0.3118	0.3022	0.3340	0.3467	0.3500	0.3545	0.3517	0.3456	0.3630	0.3538
10e1	0.2511	0.2533	0.2691	0.2531	0.2482	0.2995	0.3116	0.3147	0.3098	0.3054	0.3077	0.3213
10e2	0.2104	0.2058	0.2204	0.2286	0.2035	0.2682	0.2794	0.2775	0.2786	0.2834	0.2765	0.2647
10e3	0.1872	0.1861	0.1964	0.1803	0.1866	0.2560	0.2657	0.2775	0.2550	0.2749	0.2682	0.2654
10e4	0.1727	0.1741	0.1636	0.1687	0.1650	0.2231	0.2297	0.2329	0.2270	0.2272	0.2208	0.2359
10e5	0.1425	0.1313	0.1294	0.1229	0.1282	0.1500	0.1597	0.1594	0.1544	0.1574	0.1565	0.1541
10e6	0.1386	0.1319	0.1385	0.1274	0.1313	0.1360	0.1375	0.1365	0.1325	0.1314	0.1391	0.1392

Table 16

Means of ARI by LSS-FTC-NTR on the subject 1+5% noise using different λ_1 and λ_2, while fixing *m*=2

λ_2 \ λ_1	0	10e-4	10e-3	10e-2	10e-1	1	10e1	10e2	10e3	10e4	10e5	10e6
0	0.7254	0.7452	0.7332	0.7778	0.7948	0.7880	0.7879	0.7876	0.7997	0.7905	0.7989	0.7845
10e-4	0.7534	0.7997	0.7590	0.8253	0.8656	0.8506	0.8670	0.8615	0.8579	0.8600	0.8460	0.8406
10e-3	0.7618	0.7943	0.8057	0.8553	0.8715	0.8733	0.8780	0.8717	0.8767	0.8755	0.8781	0.8769
10e-2	0.7586	0.8079	0.8055	0.8616	**0.8906**	0.8840	0.8840	0.8878	0.8837	0.8871	0.8849	0.8838
10e-1	0.6952	0.6842	0.6858	0.6788	0.7073	0.7172	0.7190	0.7139	0.7167	0.7215	0.7187	0.7135
1	0.6621	0.6716	0.6685	0.6733	0.6787	0.6788	0.6785	0.6797	0.6742	0.6813	0.6740	0.6748
10e1	0.6514	0.6555	0.6601	0.6612	0.6671	0.6615	0.6668	0.6665	0.6668	0.6647	0.6613	0.6649
10e2	0.6599	0.6637	0.6743	0.6744	0.6702	0.6724	0.6732	0.6746	0.6737	0.6720	0.6789	0.6771
10e3	0.6621	0.6625	0.6721	0.6583	0.6612	0.6633	0.6613	0.6650	0.6696	0.6696	0.6694	0.6666
10e4	0.6567	0.6497	0.6595	0.6536	0.6576	0.6562	0.6581	0.6597	0.6537	0.6537	0.6561	0.6512
10e5	0.6495	0.6397	0.6462	0.6357	0.6378	0.6446	0.6451	0.6448	0.6408	0.6426	0.6412	0.6388
10e6	0.6456	0.6403	0.6553	0.6402	0.6409	0.6506	0.6429	0.6419	0.6489	0.6466	0.6438	0.6439

5. Conclusion

In this study, we have addressed the problem of medical image segmentation with insufficient and noisy samples, and proposed LSS-FTC-NTR model for leveraging source knowledge to improve the segmentation performance of target domain. We explore the negative-transfer-resistant mechanism to reinforce the influence of positive transfer and reduce, or even eliminate, the negative transfer. In particular, we find a shared latent space based on the idea of MMD, in which the mapped data distributions of source domain and target domain are close to each other. The experiments focus on noisy brain CT images. The experimental results show that with insufficient and noisy medical images, it is possible to build an efficient segmentation model with the help of medical images from the related scenarios. Future work will extend our algorithm to other medical image segmentation applications. We will extend the framework so as to apply various clustering algorithms in order to obtain more satisfactory medical image segmentation results. We will also study how many images in the source domain can be considered sufficient, and how to select the important images to further improve the transfer. In addition, how to speed up LSS-FTC-NTR is worthy to be studied in the future.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grants 61806026, 61702225, 61772241, and 61711540041, in part by the Natural Science Foundation of Jiangsu Province under Grants BK20180956, BK20161268 and BK20160187, in part by the Fundamental Re-search Funds for the Central Universities under Grants JUSRP51614A and JUSRP11737, in part by 2016 Qinglan Project of Jiangsu Province, and in part by Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127.

REFERENCES

[1] A. Ribbens, J. Hermans, F. Maes, D. Vandermeulen, & P. Suetens. "Unsupervised segmentation, clustering, and groupwise registration of heterogeneous populations of brain MR images." *IEEE Trans. Medical Imaging*, vol. 33, no. 2, pp. 201-224, Jun. 2013.

[2] Y. Ren, K. Hu, X. Dai, L. Pan, S. C. Hoi, & Z. Xu. "Semi-supervised deep embedded clustering. "*Neurocomputing*, vol. 325, no. 1, pp. 121-130, Jan. 2019.

[3] V. Borges, F. D. O. M. Cristina, T. Silva, A. Vieira, & B. Hamann. "Region growing for segmenting green microalgae images." *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 15, no. 1, pp. 257-270, Oct. 2018.

[4] A. Raj, A. K. Tiwari, & M. G. Martini. "Fundus image quality assessment: survey, challenges, and future scope." *IET Image Processing*, vol. 13, no. 8, pp. 1-14, Apr. 2019.

[5] P. Poudel, A. Illanes, E. J. Ataide, N. Esmaeili, S. Balakrishnan, & M. Friebe. "Thyroid Ultrasound texture classification using autoregressive features in conjunction with machine learning approaches." *IEEE Access*, vol. 7, pp. 79354-79365, Jun. 2019.

[6] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, ... & T. Vercauteren. "Interactive medical image segmentation using deep learning with image-specific fine tuning." *IEEE Trans. Medical Imaging*, vol. 37, no. 7, pp. 1562-1573, Jan. 2018.

[7] S. Wazarkar, B. N. Keshavamurthy, & A. Hussain. "Region-based segmentation of social images using soft KNN algorithm." *Procedia Computer Science*, vol. 125, no. 1, pp. 93-98, Jan. 2018.

[8] M. D. Steenwijk, P. J. Pouwels, M. Daams, J. W. van Dalen, M. W. Caan, E. Richard, ... & H. Vrenken. "Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs)." *NeuroImage: Clinical*, vol. 3, pp. 462-469, Oct. 2013.

[9] A. Feng-Ping, & L. Zhi-Wen. "Medical image segmentation algorithm based on feedback mechanism convolutional neural network." *Biomedical Signal Processing and Control*, vol. 53, pp. 101589, Aug. 2019.

[10] X. Y. Wang, T. Wang, & J. Bu. "Color image segmentation using pixel wise support vector machine classification." *Pattern Recognition*, vol. 44, no. 4, pp. 777-787, Apr. 2011.

[11] J. Enguehard, P. O'Halloran, & A. Gholipour. "Semi-supervised learning with deep embedded clustering for image classification and segmentation." *IEEE Access*, vol. 7, pp. 11093-11104, Jan. 2019.

[12] Z. Guo, X. Li, H. Huang, N. Guo, & Q. Li. "Deep learning-based image segmentation on multi-modal medical imaging." *IEEE Trans. Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 162-169, Feb. 2019.

[13] Y. Zhou, W. Huang, P. Dong, Y. Xia, & S. Wang. "D-UNET: a dimension-fusion U shape network for chronic stroke lesion segmentation." *IEEE/ACM Trans. Computational Biology and Bioinformatics*, In Press, DOI: 10.1109/TCBB.2019.2939522.

[14] J. C. Bezdek, R. Ehrlich, & W. Full. "FCM: The fuzzy c-means clustering algorithm." *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191-203, Dec. 1984.

[15] N. Dhanachandra, & Y. J. Chanu. "A survey on image segmentation methods using clustering techniques." *European Journal of Engineering Research and Science*, vol. 2, no. 1, pp. 15-20, Jan. 2017.

[16] N. M. Portela, G. D. Cavalcanti, & T. I. Ren. "Semi-supervised clustering for MR brain image segmentation." *Expert Systems with Applications*, vol. 41, no. 4, pp. 1492-1497, Mar. 2014.

[17] A. Ortiz, J. M. Gorriz, J. Ramirez, & D. Salas-Gonzalez. "Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering." *Information Sciences*, vol. 262, no. 20, pp. 117-136, Mar, 2014.

[18] S. Saha, A. K. Alok, & A. Ekbal. "Brain image segmentation using semi-supervised clustering." *Expert Systems with Applications*, vol. 52, no. 15, pp. 50-63, Jun. 2016.

[19] E. Abdel-Maksoud, M. Elmogy, & R. Al-Awadi. "Brain tumor segmentation based on a hybrid clustering technique." *Egyptian Informatics Journal*, vol. 16, no. 1, pp. 71-81, Mar. 2015.

[20] X. Wang, B. Qian, & I. Davidson. "On constrained spectral clustering and its applications." *Data Mining and Knowledge Discovery*, vol. 28, no. 1, pp. 1-30, Jan. 2014.

[21] P. Qian, Y.Jiang, Z. Deng, L. Hu, S. Sun, S. Wang, & R. F. Muzic. "Cluster prototypes and fuzzy memberships jointly leveraged cross-domain maximum entropy clustering." *IEEE Trans. cybernetics*, vol. 46, no. 1, pp. 181-193, Feb. 2015.

[22] H. Venkateswara, S. Chakraborty, & S. Panchanathan. "Deep-learning systems for domain adaptation in computer vision: Learning transferable feature representations." *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 117-129, Nov. 2017.

[23] L. Zhang, L. Zhang, D. Tao, & X. Huang. "Sparse transfer manifold embedding for hyperspectral target detection." *IEEE Trans. Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 1030-1043, Mar. 2013.

[24] C. Williams. "Transfer in context: Replication and adaptation in knowledge transfer relationships." *Strategic management journal*, vol. 28, no. 9, pp. 867-899, May. 2007.

[25] W. Jiang, & F. L. Chung. "Transfer spectral clustering." *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Berlin, Heidelberg, 2012, pp. 789-803.

[26] Z. Deng, Y. Jiang, F. L. Chung, H. Ishibuchi, K. S. Choi, & S. Wang. "Transfer prototype-based fuzzy clustering." *IEEE Trans. Fuzzy Systems*, vol. 24, no. 5, pp. 1210-1232, Dec. 2015.

[27] L. Duan, I. W. Tsang, & D. Xu. "Domain transfer multiple kernel learning." *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 465-479, Jan. 2012.

[28] P. Qian, S. Sun, Y. Jiang, K. H. Su, T. Ni, S. Wang, & R. F. Muzic Jr. "Cross-domain, soft-partition clustering with diversity measure and knowledge reference." *Pattern recognition*, vol. 50, no. 2, pp. 155-177, Feb. 2016.

[29] P. Qian, K. Zhao, Y. Jiang, K. H. Su, Z. Deng, S. Wang, & R. F. Muzic Jr. "Knowledge-leveraged transfer fuzzy C-Means for texture image segmentation with self-adaptive cluster prototype matching." *Knowledge-based systems*, vol. 130, no. 8, pp. 33-50, May. 2017.

[30] A. Charuvaka, & H. Rangwala. "Classifying protein sequences using regularized multi-task learning." *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 11, no. 6, pp. 1087-1098. Nov. 2014.

[31] W. Gu, Z. Zhang, X. Xie, & Y. He. "An improved muti-task learning algorithm for analyzing cancer survival data." *IEEE/ACM Trans. Computational Biology and Bioinformatics, In Press*. DOI: 10.1109/TCBB.2019.2920770

[32] S. F. Hussain, & M. Ramazan, "Biclustering of human cancer microarray data using co-similarity based co-clustering." *Expert Systems with Applications*, vol. 55, no. 8, pp. 520–531, Aug. 2016.

[33] S. J. Pan, & Q. Yang. "A survey on transfer learning." *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, Oct. 2009.

[34] X. Zhang, X. Zhang, H. Liu, & X. Liu. "Multi-task clustering through instances transfer." *Neurocomputing*, vol. 251, pp. 145-155, Apr. 2017.

[35] A. Van Opbroek, H. C. Achterberg, M. W. Vernooij, & M. De Bruijne. "Transfer learning for image segmentation by combining image weighting and kernel learning." *IEEE Trans. Medical Imaging*, vol. 38, no. 1, pp. 213-224, Jul. 2018.

[36] M. T. Rosenstein, Z. Marx, L. P. Kaelbling, & T. G. Dietterich. "To transfer or not to transfer." *Proceedings of the Neural Information Processing Systems 2005 Workshop on Inductive Transfer: 10 Years Later*. Cambridge: MIT Press, 2005, pp.1-4.

[37] T. Croonenborghs, K. Driessens, & M. Bruynooghe. "Learning relational options for inductive transfer in relational reinforcement learning." *International Conference on Inductive Logic Programming*. Springer, Berlin, Heidelberg, 2007, pp. 88-97.

[38] E. Talvitie, & S. P. Singh. "An Experts Algorithm for Transfer Learning." *Proceedings of the 20th International Joint Conference on Artificial Intelligence, India, January 6-12,* 2007, pp. 1065-1070.

[39] M. T. Hooijmans, O. Dzyubachyk, K. Nehrke, P. Koken, M. J. Versluis, H. E. Kan, & P. Börnert. "Fast multistation water/fat imaging at 3T using DREAM‐based RF shimming." *Journal of Magnetic Resonance Imaging,* vol. 42, no. 1, pp. 217-223, Oct. 2014.

[40] A. Kalemis, B. M. Delattre, & S. Heinzer. "Sequential whole-body PET/MR scanner: concept, clinical use, and optimisation after two years in the clinic. The manufacturer's perspective." *Magnetic Resonance Materials in Physics, Biology and Medicine,* vol. 26, no. 1, pp. 5-23, Feb. 2013.

[41] A. van Opbroek, M. A. Ikram, M. W. Vernooij, & M. de Bruijne. "Supervised image segmentation across scanner protocols: A transfer learning approach." *International Workshop on Machine Learning in Medical Imaging*. Springer, Berlin, Heidelberg, 2012, pp. 160-167.

[42] F. Nie, D. Xu, & X. Li. "Initialization independent clustering with actively self-training method." *IEEE Trans. Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 42, no. 1, pp. 17-27, Nov. 2011.

[43] J. Liu, J. Mohammed, J. Carter, S. Ranka,, T. Kahveci, & M. Baudis. "Distance-based clustering of CGH data." *Bioinformatics,* vol. 22, no. 16, pp. 1971-1978, Aug. 2006.

Yizhang Jiang (M'2012–SM'19**)** received his Ph.D. from Jiangnan University in 2015. He has also been a research assistant in the Department of Computing, Hong Kong Polytechnic University, for two years. His research interests include pattern recognition, intelligent computation and their applications. He is the author/co-author of more than 40 research papers in international/national journals, including IEEE Trans. on Fuzzy Systems, IEEE Trans. on Neural Networks and Learning Systems, IEEE Trans. on Cybernetics, and Information Sciences. He has served as a reviewer or co-reviewer of several international conferences and journals, such as ICDM, TKDE, TFS, TNNLS, Pattern Recognition, Neurocomputing, and Neural Computing & Applications. He has also served as a leader guest editor or guest editor of several international journals, such as journal of Ambient Intelligence and Humanized Computing and journal of cloud computing.

Xiaoqing Gu received his Ph.D. from Jiangnan University in 2017. She is currently a Lecturer in the School of Information Science and Engineering, Changzhou University, Changzhou, China. She has published over twenty papers in international/national authoritative journals. Her current research interests include pattern recognition and machine learning.

Dongrui Wu (S'05–M'09–SM'14) received the BE degree in automatic control from the University of Science and Technology of China in 2003, the ME degree in electrical engineering from the National University of Singapore in 2005, and the PhD degree in electrical engineering from the University of Southern California in 2009. He is now Professor in the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, and Deputy Director of the Key Laboratory of Image Processing and Intelligent Control, Ministry of Education. His research interests include affective computing, brain-computer interface, computational intelligence, and machine learning. He has more than 110 publications, including a book entitled Perceptual Computing (Wiley-IEEE Press, 2010) and 30 IEEE Transactions papers. Dr. Wu received the IEEE Computational Intelligence Society Outstanding PhD Dissertation Award in 2012, the IEEE TRANSACTIONS ON FUZZY SYSTEMS Outstanding Paper Award in 2014, the NAFIPS Early Career Award in 2014, the IEEE Systems, Man and Cybernetics (SMC) Society Early Career Award in 2017, and the IEEE SMC Society Best Associate Editor Award in 2018. He was also a finalist of another three Best Paper Awards. He was/is an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS (2011-2018), the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS (2014-), the IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE (2017-), and the IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING (2019-).

Wenlong Hang received his Ph.D. from Jiangnan University in 2017. He is currently a Lecturer in the School of Computer Science and Technology, Nanjing Tech University, Nanjing, China. He has published over ten papers in international/national authoritative journals. Her current research interests include pattern recognition and machine learning.

Jing Xue received her MD (Doctor of Medicine) degree in 2019 at the Department of Nephrology, Nanjing Medical University, China. She received her MM degree from the Department of Nephrology, Nanjing Medical University, China. Her research area includes machine learning and its applications on medicine. She is a Doctor at the Affiliated Wuxi People's Hospital of Nanjing Medical University, China.

Shi Qiu received the M.S. degree at Xidian University, Xi'an, P. R. China, in 2012, the Ph.D. degree at University of Chinese Academy of Sciences, Xi'an, P. R. China, in 2016. He is associate research fellow with Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, P. R. China. His research interests include image restoration, sparse representation, and object detection.

Chin-Teng Lin (S'88–M'91–SM'99–F'05) received the B.S. degree in control engineering from the National Chiao-Tung University (NCTU), Hsinchu, Taiwan, in 1986, and the Master's and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively. He is currently a Distinguished Professor of the Faculty of Engineering and Information Technology, University of Technology Sydney. He also holds an Honorary Chair Professorship of Electrical and Computer Engineering, NCTU, International Faculty of University of California at San-Diego, La Jolla, CA, USA, and an Honorary Professorship of the University of Nottingham, Nottingham, U.K. He is the coauthor of Neural Fuzzy Systems (Prentice-Hall, 1996), and the author of Neural Fuzzy Control Systems with Structure and Parameter Learning (World Scientific, 1994). He has published more than 220 journal papers (Total Citation: 18 236, H-index: 57, i10-index: 295) and holds 97 patents in the areas of computational intelligence, fuzzy neural networks, natural cognition, brain–computer interface, intelligent system, multimedia information processing, machine learning, robotics, and intelligent sensing and control, including approximately 108 IEEE journal papers. Dr. Lin was elevated to an IEEE Fellow for his contributions to biologically inspired information systems in 2005 and was elevated to an International Fuzzy Systems Association Fellow in 2012. He received the Merit National Science Council Research Fellow Award, Taiwan in 2009, the Outstanding Electrical and Computer Engineer, Purdue University in 2011, the Outstanding Achievement Award by Asia Pacific Neural Network Assembly in 2013, and the IEEE Fuzzy Systems Pioneer Award in 2016. He served as the Deputy Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II in 2006–2008 and as the Editor in-Chief of the IEEE Transactions on Fuzzy Systems from 2011 to 2016. He also served on the Board of Governors of the IEEE Systems, Man, Cybernetics Society during 2003–2005, the IEEE Circuits and Systems (CAS) Society during 2005–2008, and the IEEE Computational Intelligence Society (CIS) during 2008–2010. He was the Chair of the IEEE Taipei Section during 2009–2010. He was a Distinguished Lecturer of the IEEE CAS Society from 2003 to 2005 and the CIS Society from 2015 to 2017. He was the Program Chair of the IEEE International Conference on Systems, Man, and Cybernetics in 2005 and the General Chair of the 2011 IEEE International Conference on Fuzzy Systems.