WaveFuse: A Unified Unsupervised Framework for Image Fusion with Discrete Wavelet Transform – Supplementary Materials

Shaolei Liu[1,2], Manning Wang[1,2] *, and Zhijian Song[1,2]*

[1] Digital Medical Research Center, School of Basic Medical Science, Fudan University, China
[2] Shanghai Key Laboratory of Medical Image Computing and Computer Assited Intervention
slliu, mnwang, zjsong@fudan.edu.cn

1 Fusion Rule

The selection of fusion rules largely determines the quality of fused images [4]. Existing image fusion algorithms based on deep learning usually calculate the sum of the feature maps directly, leaving the information of feature maps not fully mined.

In our method, two complementary fusion rules based on DWT are adopted for wavelet components C_k transformed by feature maps F_k, including adaptive rule based on regional energy [7] and *Average* rule [2], and the fused wavelet components are denoted as F_r and F_{l1}, respectively. In adaptive rule based on regional energy, different fusion rules are employed for different frequency components, that is, the low-frequency components C_{kL} adopts an adaptive weighted averaging algorithm based on regional energy, and for the high-frequency components C_{kH}, the one with larger variance between C_{1H} and C_{2H} will be selected as the fused high-frequency components. Additionally, to preserve more structural information and make our fused image more natural, we apply l1-Norm based Average rule [2] to our fusion part, where both low and high frequency components are fused by the same rule to obtain global and general fused wavelet components. Therefore, in the fusion part, the extracted feature maps F_k are processed by two abovementioned rules, and the final fused wavelet components F is calculated by the weighted averaging of F_r and F_{l1}, which is defined as follows:

$$\begin{aligned} F(m,n) &= \omega_r F_r(m,n) + \omega_{l1} F_{l1}(m,n), \\ s.t.\ \ &\omega_r + \omega_{l1} = 1, \end{aligned} \tag{1}$$

* Corresponding authors.

where (m, n) denotes the corresponding position in C_k and F, and ω_r will be set as different values for different scenarios to achieve the optimal fusion performance. In our experiments, ω_r is set as 0.9. In the following part, these two fusion rules will be introduced in detail.

We first introduce adaptive rule based on regional energy. As is mentioned above, different fusion rules are adopted for low-frequency and high-frequency components. Therefore, F_r includes two parts, the fused low-frequency component F_{rL} and the fused high-frequency components F_{rH}. For high-frequency components, we employ a rule of maximizing variance, where F_{rH} is obtained from the one with larger variance between C_{1H} and C_{2H}, and the details are introduced in [8]. For low-frequency components, we employ a rule of weighted average based on regional energy, which is introduced as follows:

We use $E_l(m, n)$ (l = 1,2) to represent the energy in the 3×3 region centered at (m, n), which is defined as follows:

$$E_l(m,n) = \sum_{m'=-1}^{1} \sum_{n'=-1}^{1} \omega(m+m', n+n')[C_l(m+m', n+n')]^2, \tag{2}$$

$$\omega = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \tag{5a}$$

where ω means weighted coefficients and C_l represents wavelet frequency components. The matching degree M_{12} of C_{1L} and C_{2L} is defined as follows:

$$M_{12}(m,n) = \frac{2 \sum_{m'=-1}^{1} \sum_{n'=-1}^{1} \omega(\Delta m, \Delta n) C_1(\Delta m, \Delta n) C_2(\Delta m, \Delta n)}{E_1(m,n) + E_2(m,n)}, \tag{3}$$

where $\Delta m = m + m'$,$\Delta n = n + n'$. A threshold T on $M_{12}(m, n)$ is used to determine how the pixel at (m,n) is fused, and T is set as 0.8 in our network. When $M_{12}(m, n)$ is smaller than T, it means that the energy of the two feature maps in this local region is greatly discriminative. In this way, the central pixel of the region with the larger energy value will be selected as the central pixel of the feature map F_{rL}, which is calculated as follows:

$$\begin{cases} F_{rL}(m,n) = C_1(m,n), & if\ E_1(m,n) \geq E_2(m,n). \\ F_{rL}(m,n) = C_2(m,n), & if\ E_1(m,n) < E_2(m,n). \end{cases} \tag{4}$$

On the contrary, when $M_{12}(m, n)$ is greater than or equal to T, it means the two feature maps have similar energy in this local region. Consequently, a weighted fusion rule [7] is used to determine the central pixel of the feature map F_{rL}, and it is defined as follows:

$$\begin{cases} F_{rL}(m,n) = \omega_{max}(m,n) C_1(m,n) + \omega_{min}(m,n) C_2(m,n), \\ \qquad\qquad\qquad\qquad\qquad\qquad if\ E_1(m,n) \geq E_2(m,n). \\ F_{rL}(m,n) = \omega_{min}(m,n) C_1(m,n) + \omega_{max}(m,n) C_2(m,n), \\ \qquad\qquad\qquad\qquad\qquad\qquad if\ E_1(m,n) < E_2(m,n). \end{cases} \tag{5}$$

$$\omega_{max}(m,n) = \frac{1}{2} - \frac{1}{2}[\frac{1 - M_{12}(m,n)}{1 - T}], \tag{8a}$$

$$\omega_{min}(m,n) = 1 - \omega_{max}(m,n). \tag{8b}$$

Then, the feature map F_{l1} generated by *l1-Norm* rule is denoted as follows:

$$F_{L1}(m,n) = \sum_{l=1}^{2} \omega_l(m,n) \times C_l(m,n), \tag{6}$$

$$\omega_l(m,n) = \frac{\hat{C}_l(m,n)}{\sum_{\xi=1}^{2} \hat{C}_\xi(m,n)}, \tag{9a}$$

$$\hat{C}_l(m,n) = \frac{\sum_{a=-r}^{r} \sum_{b=-r}^{r} ||C_l(m+a, n+b)||_1}{(2r+1)^2}, \tag{9b}$$

where r means the block size and r is set as 1 according to [2].

2 Experimental Results and Analysis

Table 1. Quantitative comparison of WaveFuse with different training datasets. The metrics of MINI are obtained by the average of MINI1-MINI3. Red ones are the best metrics. For all metrics, larger is better.

	dataset	EN	CE	FMI_pixel	FMI_dct	FMI_w	Q^{NICE}	$Q^{AB/F}$	VARI	MS-SSIM
ME	**coco**	**6.9224**	4.7309	0.8661	**0.4941**	**0.5265**	0.8240	0.5149	**39.8423**	0.9582
	mini1	6.8838	4.8488	0.8488	0.2544	0.2921	0.8159	0.4469	38.8022	0.9560
	mini2	6.8941	4.7463	**0.8668**	0.4919	0.5212	**0.8242**	**0.5184**	39.3760	**0.9611**
	mini3	6.9071	**4.8689**	0.8613	0.3528	0.3820	0.8189	0.4927	39.1288	0.9590
MED	**coco**	5.2814	7.8243	**0.8650**	**0.4052**	**0.3848**	**0.8094**	0.3265	67.0572	0.9007
	mini1	5.5875	7.1893	0.8518	0.2113	0.2333	0.8080	0.2950	**69.4813**	0.8966
	mini2	5.3457	**8.1940**	0.8647	0.4001	0.3841	**0.8094**	**0.3270**	68.2492	0.9000
	mini3	**5.5899**	6.7014	0.8595	0.2776	0.2785	0.8084	0.3110	68.6736	**0.9025**
MF	**coco**	7.3681	**0.3483**	**0.8912**	**0.5041**	**0.5619**	0.8368	**0.7743**	51.7256	**0.9896**
	mini1	7.3822	0.3214	0.8672	0.2711	0.3326	0.8291	0.6841	51.3683	0.9812
	mini2	7.3772	0.3407	0.8911	0.5014	0.5593	**0.8371**	0.7737	52.1674	0.9883
	mini3	**7.3879**	0.3215	0.8822	0.3593	0.4156	0.8314	0.7347	**51.7589**	0.9853
IV	**coco**	**6.7805**	1.4595	**0.8891**	**0.3853**	**0.4122**	0.8057	0.5449	**34.2832**	**0.9268**
	mini1	6.7640	1.4662	0.8777	0.1970	0.2476	0.8055	0.5040	34.1604	0.9204
	mini2	6.7409	**1.5072**	0.8877	0.3809	0.4071	**0.8058**	**0.5480**	33.5035	0.9243
	mini3	6.7520	1.4042	0.8842	0.2675	0.3115	0.8056	0.5333	33.7721	0.9258

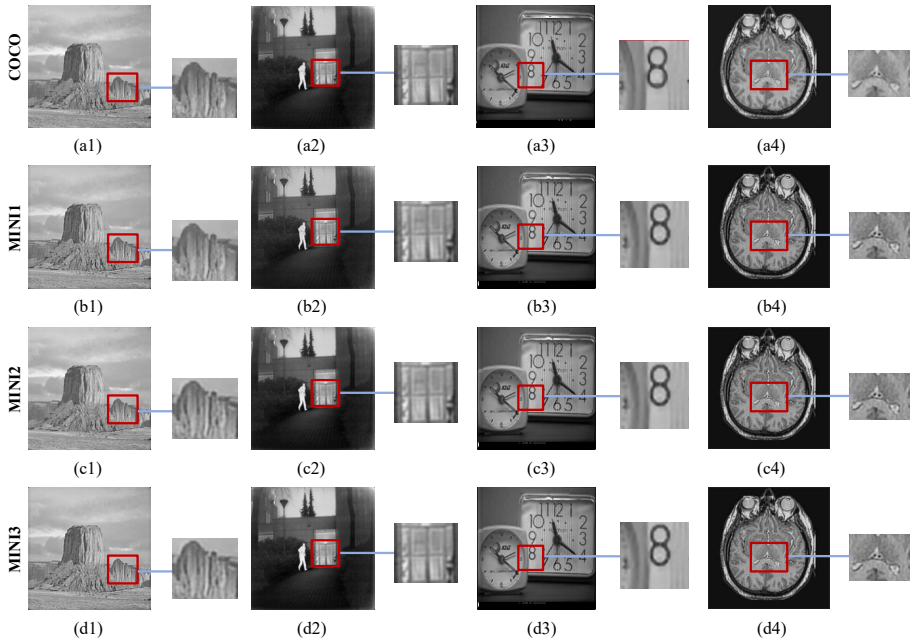

Fig. 1. Fusion results with different training datasets:MINI1-MINI3, each of which contains 0.5%, 1% and 2% images respectively chosen randomly from COCO.

2.1 Comparison of Using Different Training Dataset

In order to further demonstrate the effectiveness and robustness of our network, we conducted experiments on another three different training minisets: MINI1-MINI3, each of which contains 0.5%, 1% and 2% images respectively chosen randomly from COCO, and the fusion results are shown in Fig. 1 and Table 1. We compared the fusion results of WaveFuse on COCO and MINI1-MINI3. The same three sets of images and ω_{RE} values were chosen for testing process. The fusion performance was compared and analyzed by the averaged fusion quality metrics. From the fusion results in Fig. 1, no obvious visual difference can be found among the fused images in WaveFuse when different training sets are chosen. Then objective metrics are employed to evaluate the fusion performance.

The fusion metrics of WaveFuse trained with the different training datasets are shown in Table 1. We can learn that even trained with limited images, WaveFuse can still achieve promising results compared with that trained with COCO dataset. In WaveFuse, higher performance is even achieved by training on minisets. Furthermore, from Table 2, we can observe that WaveFuse is trained on minisets within one hour, where the GPU memory utilization is just 4085 MB, so it can be trained with lower computational cost. Accordingly, we can learn that our proposed network is robust both to the size of the training dataset and to the selection of training images.

Table 2. Comparison of training efficiency in WaveFuse with different datasets.

	Time(h)	GPU(MB)
COCO	7.78	17345
MINI1	0.13	4085
MINI2	0.18	4085
MINI3	0.44	4085

2.2 Ablation Studies

• **DWT-based feature fusion** In this section, we attempt to explain why DWT-based feature fusion module can improve fusion performance. DWT has been a poweful multi-sacle analysis tool in signal and image processing since it was proposed. DWT transforms the images into different low and high frequencies, where low frequencies represent contour and edge information and high frequencies represent detailed texture information[1]. In this way, DWT-based fusion methods first transform the images into low and high frequencies, and then fuse them in the wavelet domain, achieving promising fusion results. Inspired by DWT methods, we apply DWT-based fusion module to deep feature fusion extracted by deep-learning models, so as to fully utilize the information contained in deep features. We conducted the ablation study about DWT-based

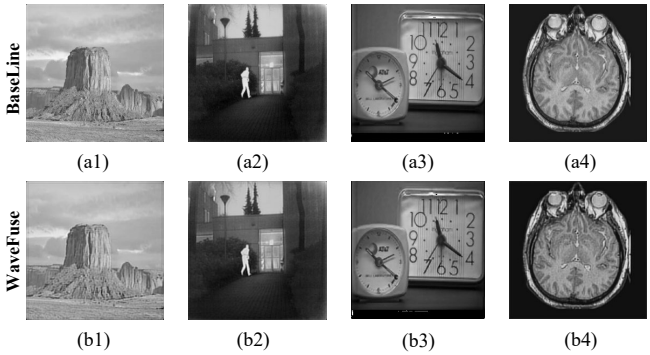

Fig. 2. Subjective ablation study on the DWT-based feature fusion module.

feature fusion module, and the results is shown in Table 3. As we can see, when we apply the module, the fusion performance is indeed improved largely.

Table 3. Ablation study on the DWT-based feature fusion module. Red ones are the best results. For all metrics, larger is better.

		EN	CE	FMI_pixel	FMI_dct	FMI_w	Q^{NICE}	$Q^{A/BF}$	VARI	MS-SSIM
ME	BaseLine	6.8230	**4.9708**	0.8473	0.2500	0.2875	0.8153	0.4451	37.5291	0.9566
	WaveFuse	**6.9224**	4.7309	**0.8661**	**0.4941**	**0.5265**	**0.8240**	**0.5149**	**39.8423**	**0.9582**
MED	BaseLine	**5.5631**	**7.8961**	0.8505	0.2098	0.2274	0.8078	0.2911	66.7267	0.8970
	WaveFuse	5.2814	7.8243	**0.8650**	**0.4052**	**0.3848**	**0.8094**	**0.3265**	**67.0572**	**0.9007**
MF	BaseLine	**7.3759**	0.3276	0.8663	0.2686	0.3314	0.8290	0.6821	51.0500	0.9812
	WaveFuse	7.3679	**0.3498**	**0.8909**	**0.5028**	**0.5609**	**0.8368**	**0.7746**	**51.6778**	**0.9889**
IV	baseline	6.7367	1.4337	0.8776	0.1919	0.2443	0.8054	0.5036	33.1956	0.9182
	WaveFuse	**6.7805**	**1.4595**	**0.8891**	**0.3853**	**0.4122**	**0.8057**	**0.5449**	**34.2832**	**0.9268**

• Experiments on Different Wavelet Decomposition Layers and Different Wavelet Bases

In wavelet transform, the number of decomposition layers and the selection of different wavelet bases could exert great impacts on the effectiveness of wavelet transform. In the following experiments with COCO dataset, different wavelet decomposition layers and bases are selected for further optimization on our proposed method. We chose decomposition layers from 1 to 4, and wavelet base was set as *db1* [3] in this experiment. From Table 4 and Fig. 3, we can clearly see that when the layer is set as 2, the best fusion performance is achieved.

For the comparison of using different wavelet bases, we set the decomposition layer as 3, and four bases including *sym2* [5], *sym3* [5], *db1* [3] and *rbio6.8* [6] were chosen. From a subjective point of view in Fig. 4, we find it difficult to distinguish which wavelet base achieves better fusion performance. Combined with the objective evaluation metrics in Table. 5, the fusion quality of wavelet base *db1* is the highest in the four image fusion scenarios. Through the above two experiments, we can further improve our proposed method by selecting the

Table 4. Quantitative comparison with different wavelet decomposition layers in WaveFuse. Red ones are the best results. For all metrics, larger is better.

	layer	EN	CE	FMI_pixel	FMI_dct	FMI_w	Q^{NICE}	$Q^{AB/F}$	VARI	MS-SSIM
ME	**1**	6.8983	4.8336	0.8482	0.2544	0.2930	0.8161	0.4487	39.1743	0.9561
	2	**6.9224**	4.7309	**0.8661**	**0.4941**	**0.5265**	**0.8240**	**0.5149**	**39.8423**	**0.9582**
	3	6.8721	4.8650	0.8488	0.2538	0.2914	0.8157	0.4466	38.5617	0.9564
	4	6.8662	**4.8678**	0.8491	0.2539	0.2908	0.8155	0.4470	38.3903	0.9570
MED	**1**	5.6018	6.7879	0.8521	0.2112	0.2332	0.8080	0.2939	70.2801	0.8964
	2	5.2814	**7.8243**	**0.8650**	**0.4052**	**0.3848**	**0.8094**	**0.3265**	67.0572	0.9007
	3	5.6085	7.1616	0.8516	0.2102	0.2304	0.8080	0.2969	**68.9429**	0.8992
	4	**5.7283**	6.7363	0.8517	0.2099	0.2227	0.8078	0.2933	68.4144	**0.8999**
MF	**1**	**7.3856**	0.3239	0.8667	0.2704	0.3317	0.8292	0.6836	51.4828	0.9805
	2	7.3681	**0.3483**	**0.8912**	**0.5041**	**0.5619**	**0.8368**	**0.7743**	**51.7256**	**0.9896**
	3	7.3828	0.3193	0.8668	0.2701	0.3312	0.8291	0.6829	51.3509	0.9813
	4	7.3821	0.3190	0.8668	0.2702	0.3314	0.8291	0.6828	51.3217	0.9813
IV	**1**	6.7741	1.4592	0.8774	0.1976	0.2477	0.8055	0.5024	34.4736	0.9202
	2	**6.7805**	1.4595	**0.8891**	**0.3853**	**0.4122**	**0.8057**	**0.5449**	**34.2832**	**0.9268**
	3	6.7560	**1.4695**	0.8787	0.1962	0.2469	0.8054	0.5053	33.9013	0.9229
	4	6.7550	1.4523	0.8795	0.1961	0.2470	0.8054	0.5062	33.8602	0.9253

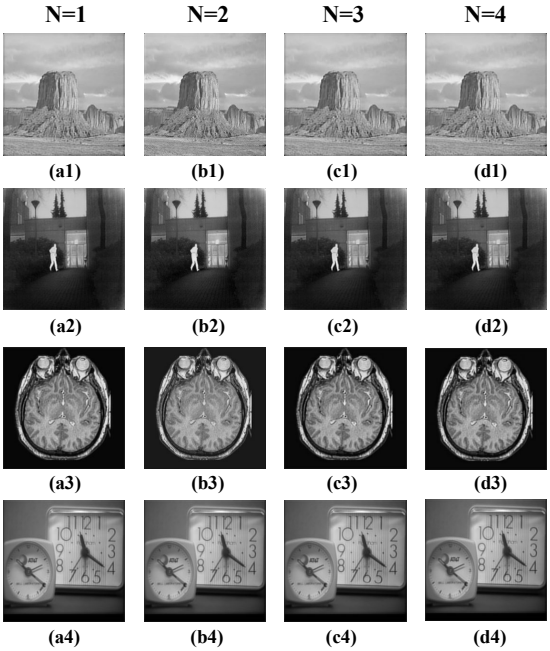

Fig. 3. Fusion results obtained by our WaveFuse with different wavelet decomposition layers.

Table 5. Quantitative comparison with different wavelet bases in WaveFuse. Red ones are the best results. For all metrics, larger is better.

	base	EN	CE	FMI_pixel	FMI_dct	FMI_w	Q^{NICE}	$Q^{AB/F}$	VARI	MS-SSIM
ME	**sym2**	6.8852	4.6731	0.8481	0.2541	0.2916	0.8159	0.4457	38.7944	0.9560
	sym3	6.8853	**4.8450**	0.8477	0.2544	0.2916	0.8159	0.4449	38.7831	0.9552
	db1	**6.9224**	4.7309	**0.8661**	**0.4941**	**0.5265**	**0.8240**	**0.5149**	**39.8423**	**0.9582**
	rbio6.8	6.8875	4.6840	0.8474	0.2544	0.2913	0.8158	0.4448	38.8210	0.9556
MED	**sym2**	5.5952	7.3196	0.8515	0.2109	0.2313	0.8080	0.2941	69.6521	0.8965
	sym3	**5.6102**	7.0149	0.8516	0.2110	0.2306	0.8080	0.2943	**69.6753**	0.8963
	db1	5.2814	**7.8243**	**0.8650**	**0.4052**	**0.3848**	**0.8094**	**0.3265**	67.0572	**0.9007**
	rbio6.8	5.4939	7.1212	0.8488	0.2185	0.2277	0.8072	0.2746	61.7255	0.8878
MF	**sym2**	7.3854	0.3216	0.8668	0.2705	0.3315	0.8291	0.6818	51.4506	0.9811
	sym3	7.3856	0.3343	0.8669	0.2704	0.3315	0.8291	0.6814	51.4693	0.9811
	db1	7.3681	**0.3483**	**0.8912**	**0.5041**	**0.5619**	**0.8368**	**0.7743**	**51.7256**	**0.9896**
	rbio6.8	**7.3863**	0.3350	0.8668	0.2705	0.3313	0.8291	0.6813	51.4698	0.9810
IV	**sym2**	6.7667	1.4625	0.8772	0.1974	0.2471	0.8055	0.5010	34.2315	0.9202
	sym3	6.7659	1.4638	0.8767	0.1979	0.2472	0.8055	0.5000	34.2216	0.9197
	db1	**6.7805**	1.4595	**0.8891**	**0.3853**	**0.4122**	**0.8057**	**0.5449**	**34.2832**	**0.9268**
	rbio6.8	6.7663	**1.4646**	0.8770	0.1978	0.2473	0.8054	0.4996	34.2271	0.9195

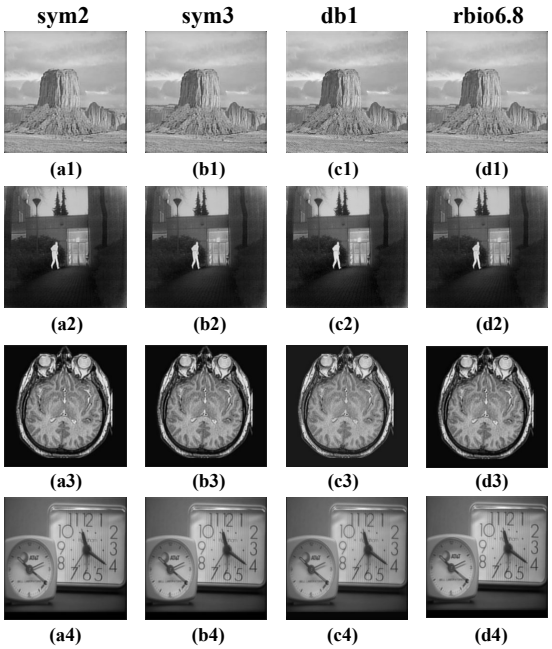

Fig. 4. Fusion results obtained by our WaveFuse with different wavelet bases.

appropriate number of decomposition layers and wavelet bases, providing a new direction for the follow-up improvement of our method.

References

1. Li, H., Manjunath, B., Mitra, S.K.: Multisensor image fusion using the wavelet transform. Graphical models and image processing **57**(3), 235–245 (1995)
2. Li, H., Wu, X.J.: Densefuse: A fusion approach to infrared and visible images. IEEE Transactions on Image Processing **28**(5), 2614–2623 (2018)
3. Lina, J.M., Mayrand, M.: Complex daubechies wavelets. Applied and Computational Harmonic Analysis **2**(3), 219–229 (1995)
4. Liu, Y., Chen, X., Peng, H., Wang, Z.: Multi-focus image fusion with a deep convolutional neural network. Information Fusion **36**, 191–207 (2017)
5. Singh, R., Vasquez, R.E., Singh, R.: Comparison of daubechies, coiflet, and symlet for edge detection. In: Visual Information Processing VI. vol. 3074, pp. 151–159. International Society for Optics and Photonics (1997)
6. Sweldens, W.: Lifting scheme: a new philosophy in biorthogonal wavelet constructions. In: Wavelet applications in signal and image processing III. vol. 2569, pp. 68–79. International Society for Optics and Photonics (1995)
7. Xiao-hua, S., Yang, G.s., Zhang, H.l.: Improved on the approach of image fusion based on region-energy. Journal of Projectiles, Rockets, Missiles and Guidance **4** (2006)
8. Zhang, B.: Study on image fusion based on different fusion rules of wavelet transform. In: 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE). vol. 3, pp. V3–649. IEEE (2010)