Ground-Truth Free Meta-Learning for Deep Compressive Sampling (Supplemental Material)

1. Network Structure

See Figure 1 for the detailed structure of our bias-adaptive unrolling CNN. Our method only adjusts the biases of all convolutional layers except for the first and last ones during adaption.

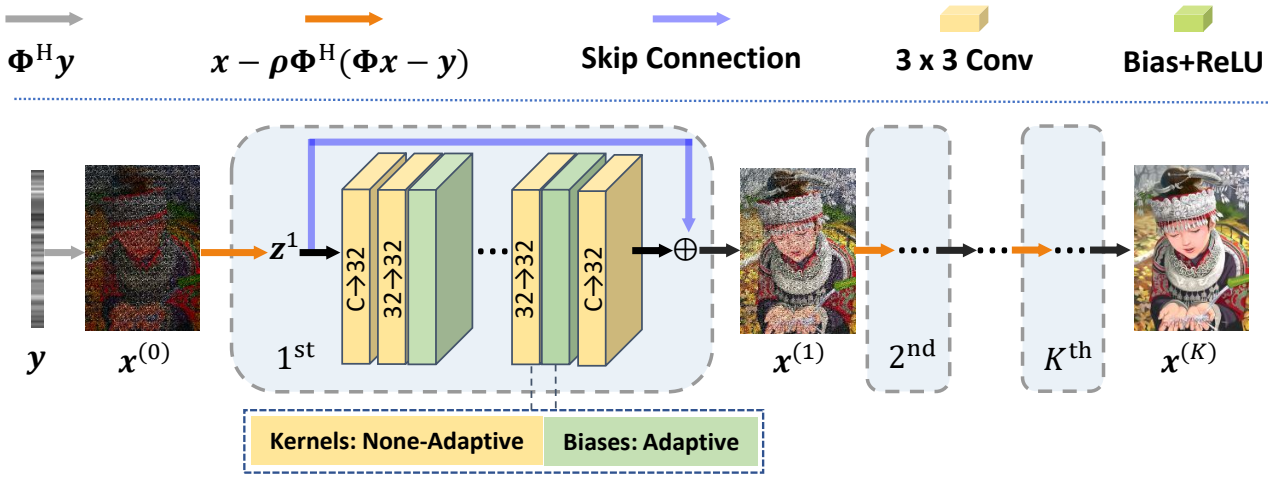


Figure 1. Illustration of our bias-adaptive unrolling CNN. It contains K blocks. Each block is composed of six convolution layers with the kernel size of 3×3, and with a ReLU activation equipped in the second to forth layers. Only the biases of all convolutional layers except for the first and last ones are adjusted during adaption.

2. Proof of Theorem 1

Proof.

$$\nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\ell^{\mathrm{iSURE}}(\boldsymbol{\omega};\boldsymbol{y},\boldsymbol{\Phi},\boldsymbol{\epsilon}') = \nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\|\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})-\boldsymbol{y}\|_2^2 + 2\sigma^2\mathrm{tr}\Big(\boldsymbol{\Phi}\frac{\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})}{\partial\boldsymbol{y}}\Big)\Big] \tag{1}$$

Let $\boldsymbol{z} = \boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})$ which is a function of $\boldsymbol{y}+\boldsymbol{\epsilon}'$. We have $\partial\boldsymbol{z}/\partial\boldsymbol{\epsilon}' = \partial\boldsymbol{z}/\partial\boldsymbol{y}$ and

$$\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\mathrm{tr}\Big(\boldsymbol{\Phi}\frac{\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})}{\partial\boldsymbol{y}}\Big)\Big] = \mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\sum_{i=1}^{M}\frac{\partial\boldsymbol{z}_i}{\partial\boldsymbol{\epsilon}_i'}\Big] = \mathbb{E}_{\boldsymbol{y}}\Big[\sum_{i=1}^{M}\int_{-\infty}^{+\infty}\frac{\partial\boldsymbol{z}_i}{\partial\boldsymbol{\epsilon}_i'}\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i'\Big], \tag{2}$$

where $\phi_\sigma(\cdot):\mathbb{R}\to\mathbb{R}$ is the 1D Gaussian p.d.f. with standard deviation σ. Using $\nabla\phi_\sigma(x) = -\frac{1}{\sigma^2}x\phi_\sigma(x)$, we have that

$$\begin{aligned}\int_{-\infty}^{+\infty}\frac{\partial\boldsymbol{z}_i}{\partial\boldsymbol{\epsilon}_i'}\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i' &= \boldsymbol{z}_i\phi_\sigma(\boldsymbol{\epsilon}_i')|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty}\boldsymbol{z}_i\nabla\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i' = \boldsymbol{z}_i\phi_\sigma(\boldsymbol{\epsilon}_i')|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty}\frac{1}{\sigma^2}\boldsymbol{z}_i\boldsymbol{\epsilon}_i'\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i' \\ &= \int_{-\infty}^{+\infty}\frac{1}{\sigma^2}\boldsymbol{z}_i\boldsymbol{\epsilon}_i'\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i = \frac{1}{\sigma^2}\mathbb{E}_{\boldsymbol{\epsilon}_i'}\boldsymbol{z}_i\boldsymbol{\epsilon}_i'.\end{aligned} \tag{3}$$

Substituting Eq. (3) back into Eq. (2), we have

$$\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\mathrm{tr}\Big(\boldsymbol{\Phi}\frac{\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})}{\partial\boldsymbol{y}}\Big)\Big]=\mathbb{E}_{\boldsymbol{y}}\Big[\sum_{i=1}^{M}\frac{1}{\sigma^2}\mathbb{E}_{\boldsymbol{\epsilon}_i'}\boldsymbol{z}_i\boldsymbol{\epsilon}_i'\Big]=\frac{1}{\sigma^2}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[(\boldsymbol{\epsilon}')^{\mathrm{H}}\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big]. \tag{4}$$

Finally, substituting Eq. (4) back into Eq. (1) gives us

$$\begin{aligned}\nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\ell^{\mathrm{iSURE}}(\boldsymbol{\omega};\boldsymbol{y},\boldsymbol{\Phi},\boldsymbol{\epsilon}')=&\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[2\boldsymbol{J}_{\boldsymbol{\omega}}^{\mathrm{H}}\Big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big)\boldsymbol{\Phi}^{\mathrm{H}}\Big(\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})-\boldsymbol{y}\Big)\Big]+2\nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[(\boldsymbol{\epsilon}')^{\mathrm{H}}\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big]\\=&\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[2\boldsymbol{J}_{\boldsymbol{\omega}}^{\mathrm{H}}\Big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big)\boldsymbol{\Phi}^{\mathrm{H}}\Big(\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})-\boldsymbol{y}\Big)\Big]+2\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\boldsymbol{J}_{\boldsymbol{\omega}}^{\mathrm{H}}\Big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big)\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\epsilon}'\Big]\\=&2\,\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\boldsymbol{J}_{\boldsymbol{\omega}}^{\mathrm{H}}\Big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big)\boldsymbol{\Phi}^{\mathrm{H}}\Big(\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})-\boldsymbol{y}+\boldsymbol{\epsilon}'\Big)\Big].\end{aligned} \tag{5}$$

The proof is done. □

Remark 1. *Our theorem and proof consider the case that both the measurements and the measurement noise are real-valued. During implementation, we treat them as real variables with double dimensions, and all expectations involving these complex-valued variables are calculated using real calculus. The proof may be extended to the complex-valued case using Wirtinger derivatives.*

3. Additional Quantitative Experiments

3.1. MR Image Reconstruction

We also use Gaussian masks of ratios $\frac{1}{5}$, $\frac{1}{4}$, $\frac{1}{3}$ for the experiments on the ADNI dataset. See Table 1 for the results. Other experimental settings are kept unchanged. It can be seen that the proposed MetaCS performs the best overall among all GT-free methods and performs competitively with the supervised methods.

Noise	CS Ratio	Non-Learning		Unsupervised	Internal		Unsupervised + Internal		Supervised	
		ZF	SparseMRI	REI	BNN	ASGLD	DDSSL	MetaCS	ADMMNet	Supervised
w/o	1/3	27.22/.71	34.93/.93	37.34/.94	37.60/.94	37.79/.94	37.81/.95	**37.87/.96**	38.22/.98	38.95/.98
	1/4	26.16/.68	32.79/.90	36.08/.94	36.10/.95	36.07/.95	36.68/**.96**	**36.71/.96**	35.94/.96	37.42/.97
	1/5	25.66/.66	31.69/.89	34.35/.92	33.81/.93	34.40/.92	35.43/**.94**	**35.52/.94**	34.81/.96	36.09/.95
w/	1/3	26.60/.65	27.91/.69	29.95/.80	29.46/.75	29.80/.76	31.59/.87	**31.76/.88**	31.04/.86	31.56/.88
	1/4	25.79/.64	27.42/.69	29.87/.79	29.20/.77	29.61/.78	30.91/.86	**31.38/.87**	30.92/.86	31.23/.87
	1/5	25.40/.63	26.97/.68	29.21/.77	29.17/.76	29.45/.75	29.39/.86	**30.95/.87**	30.81/.86	31.02/.88

Table 1. Mean PSNR(dB)/SSIM results of natural image reconstruction. **Boldfaced**: best of all compared **GT-free** methods.

3.2. Natural Image Reconstruction

We also test MetaCS with natural image reconstruction on BSD68 [3]. The results are reported in Table 2. Again, MetaCS performs the best overall among all GT-free methods and performs competitively with the supervised methods.

Noise	CS Ratio	Non-Learning	Unsupervised		Internal		Unsupervised+Internal		Supervised		
		TVAL3	LSURE	REI	BNN	ASGLD	DDSSL	MetaCS	Supervised	COAST	SSLIP
w/o	40%	29.39/.86	31.87/.90	31.79/.90	31.28/.90	31.36/.90	32.53/**.92**	**32.63/.92**	32.17/.92	33.02/.92	30.72/.88
	25%	26.48/.77	28.73/.84	28.45/.82	28.63/.84	29.51/.84	29.47/**.86**	**29.67/.86**	29.36/.85	30.07/.87	28.26/.81
	10%	22.49/.58	23.07/.65	23.11/.63	25.24/.71	25.51/.70	**26.03/.72**	25.88/**.72**	25.32/.71	26.25/.72	24.72/.66
w/	40%	26.15/.68	27.73/.77	28.05/.79	28.13/.81	28.75/.81	29.21/**.84**	**29.49/.84**	26.86/.72	28.98/.83	28.47/.83
	25%	24.75/.67	28.14/.82	28.08/.81	28.67/.84	29.35/.85	29.61/**.87**	**29.71/.87**	29.49/.86	29.37/.86	28.71/.85
	10%	22.03/.52	23.54/.60	22.37/.60	23.79/.64	24.56/.64	24.63/.67	**24.67/.68**	23.86/.60	24.02/.67	24.25/.67

Table 2. Mean PSNR(dB)/SSIM results of natural image reconstruction. **Boldfaced**: best of all compared **GT-free** methods.

3.3. Performance Using a Larger Model

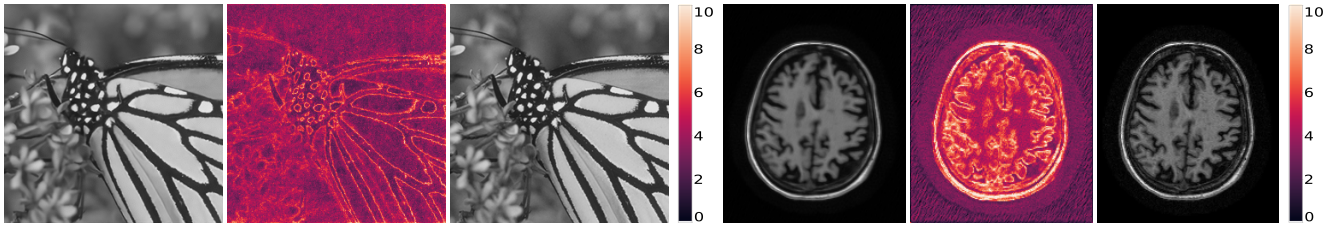
We enlarge the model size of MetaCS to 0.75M (versus 0.79M of the original DDSSL) by setting $K = 20$ while keeping other settings unchanged. See Table 3 for its results with comparison to the original DDSSL (without reducing its model but quoting results from its paper). Clearly, our MetaCS still performs better in both reconstruction accuracy and complexity.

Noise	Method	MRI - MRI150			MRI - ADNI			Natural Images			Time
		$r = 20\%$	30%	40%	r =1/5	1/4	1/3	$r = 10\%$	25%	40%	(min.)
w/o	DDSSL	36.75	38.48	41.03	33.90	34.67	36.43	27.59	33.41	37.19	7.28
	Ours	**37.15**	**39.63**	**41.26**	**34.11**	**34.86**	**36.59**	**28.06**	**33.58**	**37.36**	**1.81**
w/	DDSSL	33.82	34.56	34.90	30.49	31.08	31.71	26.12	29.84	31.94	7.28
	Ours	**34.38**	**35.27**	**35.62**	**30.62**	**31.21**	**31.84**	**26.25**	**29.89**	**31.97**	**1.81**

Table 3. PSNR results of original DDSSL and enlarged MetaCS. The time is reported on natural image reconstruction.

4. Uncertainty Quantization

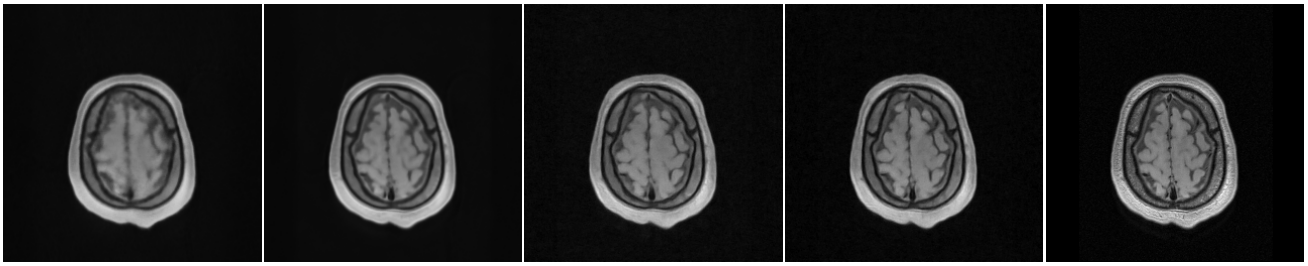
Uncertainty quantification on reconstructed images is an important feature often requested in scientific imaging and medical imaging. Recall that the ensemble inference provides a number of instances, which, as a byproduct, can be used for uncertainty quantization of CSR. Concretely, we calculate the standard deviation of $\mathcal{F}_{\boldsymbol{\omega}^*}(\boldsymbol{y}^* + \boldsymbol{\epsilon}', \boldsymbol{\Phi})$ over $\boldsymbol{\epsilon}'$ as the uncertainty map. Figure 2 visualizes such maps on two samples. We can see that flat regions have lower uncertainty while structure regions (*e.g.* edges and textures) have higher uncertainty.



Reconstructed image | Uncertainty map | GT | Reconstructed image | Uncertainty map | GT

Figure 2. Visualization of uncertainty maps of noisy CSR using 10 instances. Up: Natural image, r=40%; bottom: MR image, r=1/3.

5. Visual Comparison on Reconstructed Images



REI | ASGLD | DDSSL | MetaCS | GT

Figure 3. Visual results of MR image reconstruction with unknown noise type.

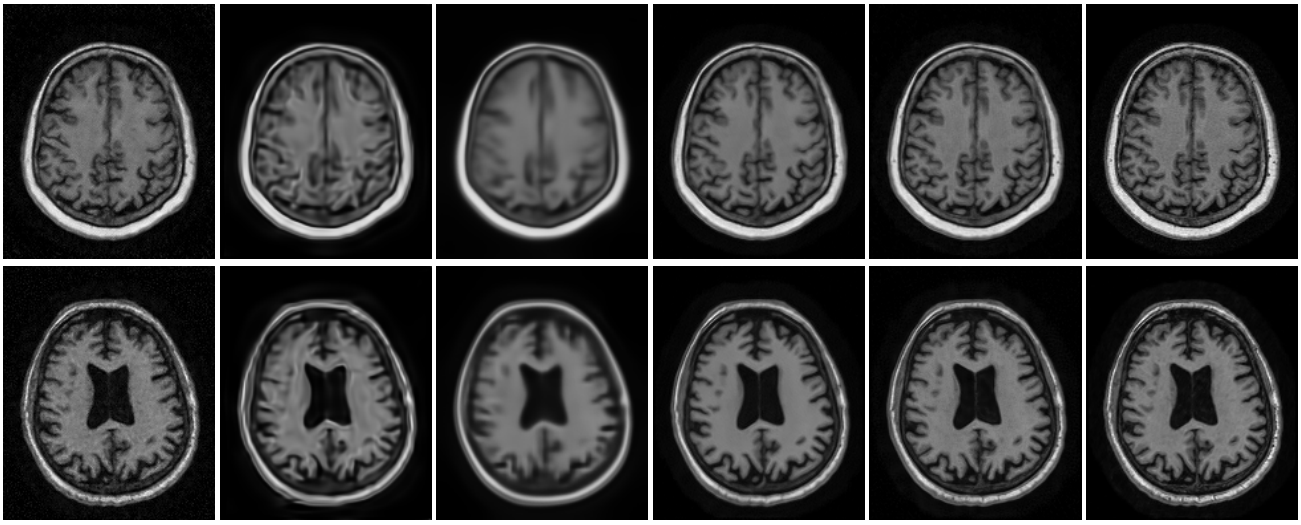


Figure 4. Visual results of MR image reconstruction with radial mask (the first row) and 2D Gaussian mask (the second row) of sampling ratio $1/4$ in noisy setting.

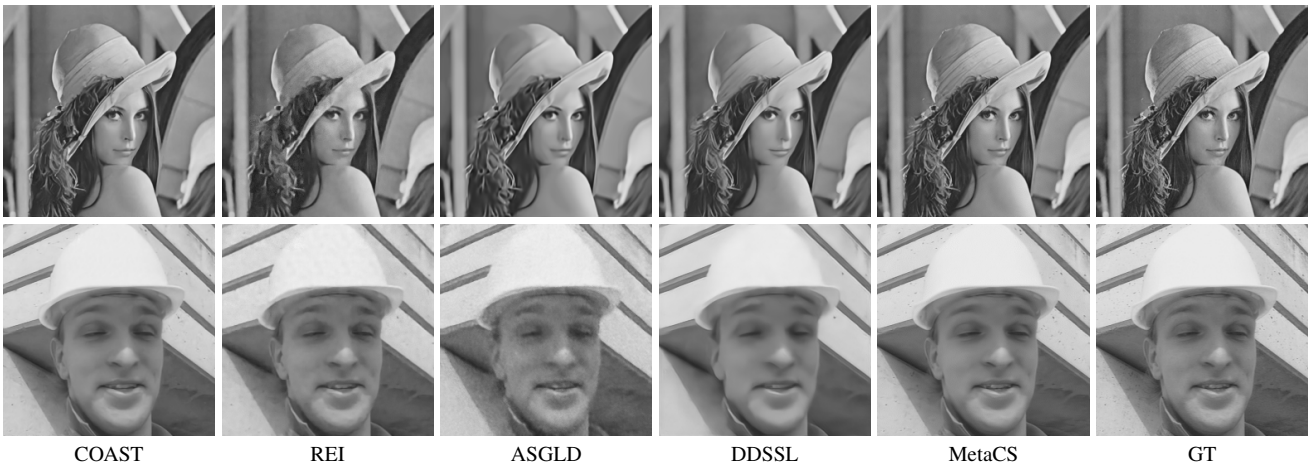


Figure 5. Visual results of natural image reconstruction from Gaussian measurements of sampling ratio 40% in noisy setting (second row).

References

[1] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619, 1984.

[2] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael A. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transacitons on Image Processing*, 26:4509–4522, 2017.

[3] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of International Conference on Computer Vision*, volume 2, pages 416–423. IEEE, 2001. 2

[4] C McCollough. Tu-fg-207a-04: Overview of the low dose ct grand challenge. *Medical Physics*, 43(6Part35):3759–3760, 2016.