

## Sparse Models vs. Representations

- Better encoding or “Representation” schemes perform better for a given best sparse model called the “Dictionary”.
- Recent results show that a simple k-means based dictionary or just using random examples from the training data perform well [1].
- **Weak Sparse Models:**
  - Weaker models from small subsets of training data.
  - *Can we combine multiple weak sparse models to improve performance?*
- **Weak Representations:**
  - 1-sparse representations / Correlate-and-Maximize.
  - *Can we combine multiple weak representations from weak sparse models to improve performance?*
- **Ensemble Approach:**
  - For the data  $x$ , compute  $L$  sparse approximations, each using a different dictionary  $D_{l+1}$

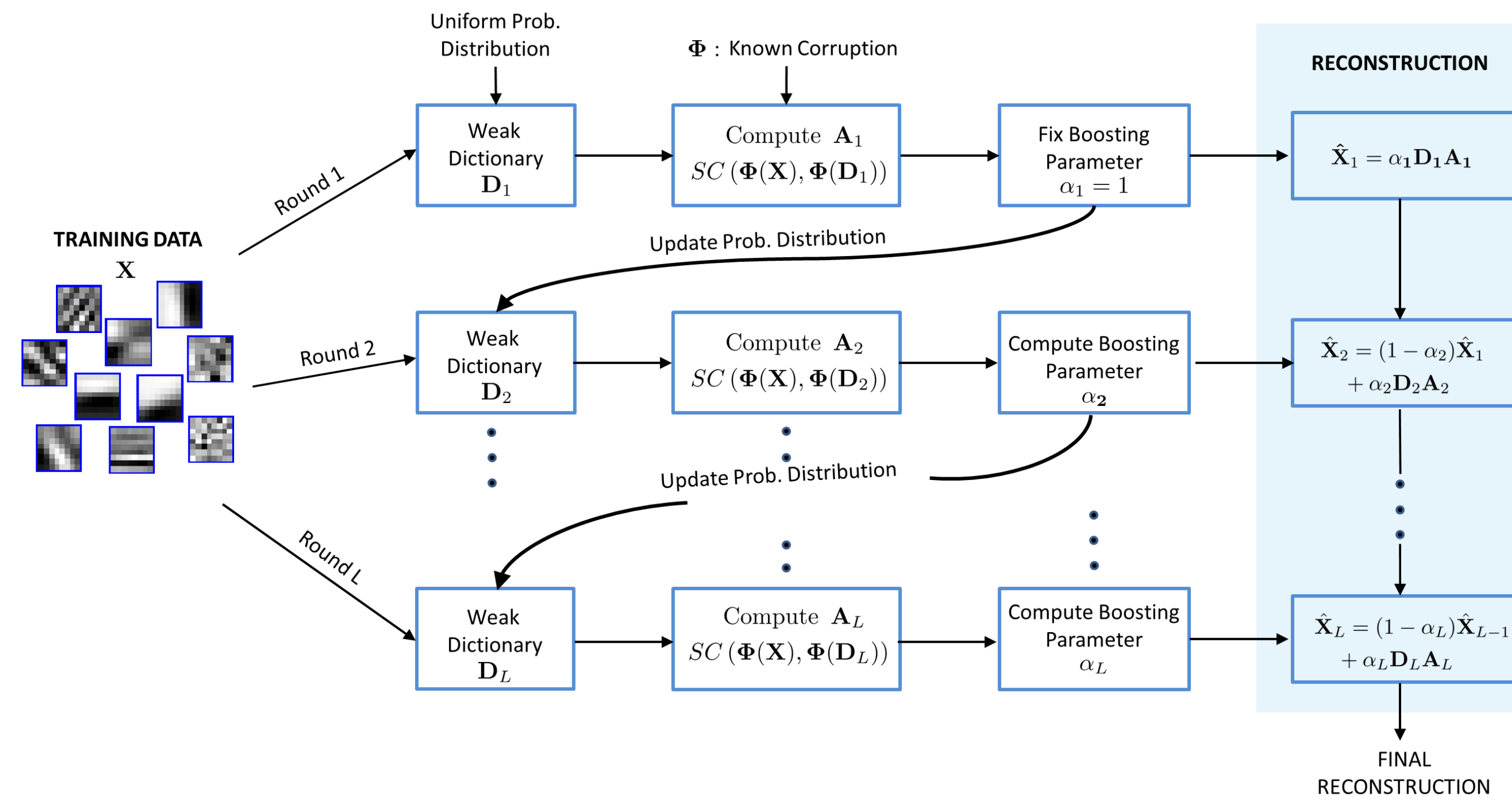
$$\{D_1 a_1, D_2 a_2, \dots, D_L a_L\}$$

- The ensemble approximation of  $x$  is the linear combination

$$\sum_{l=1}^L \beta_l D_l a_l \quad \text{Parameters of the ensemble}$$

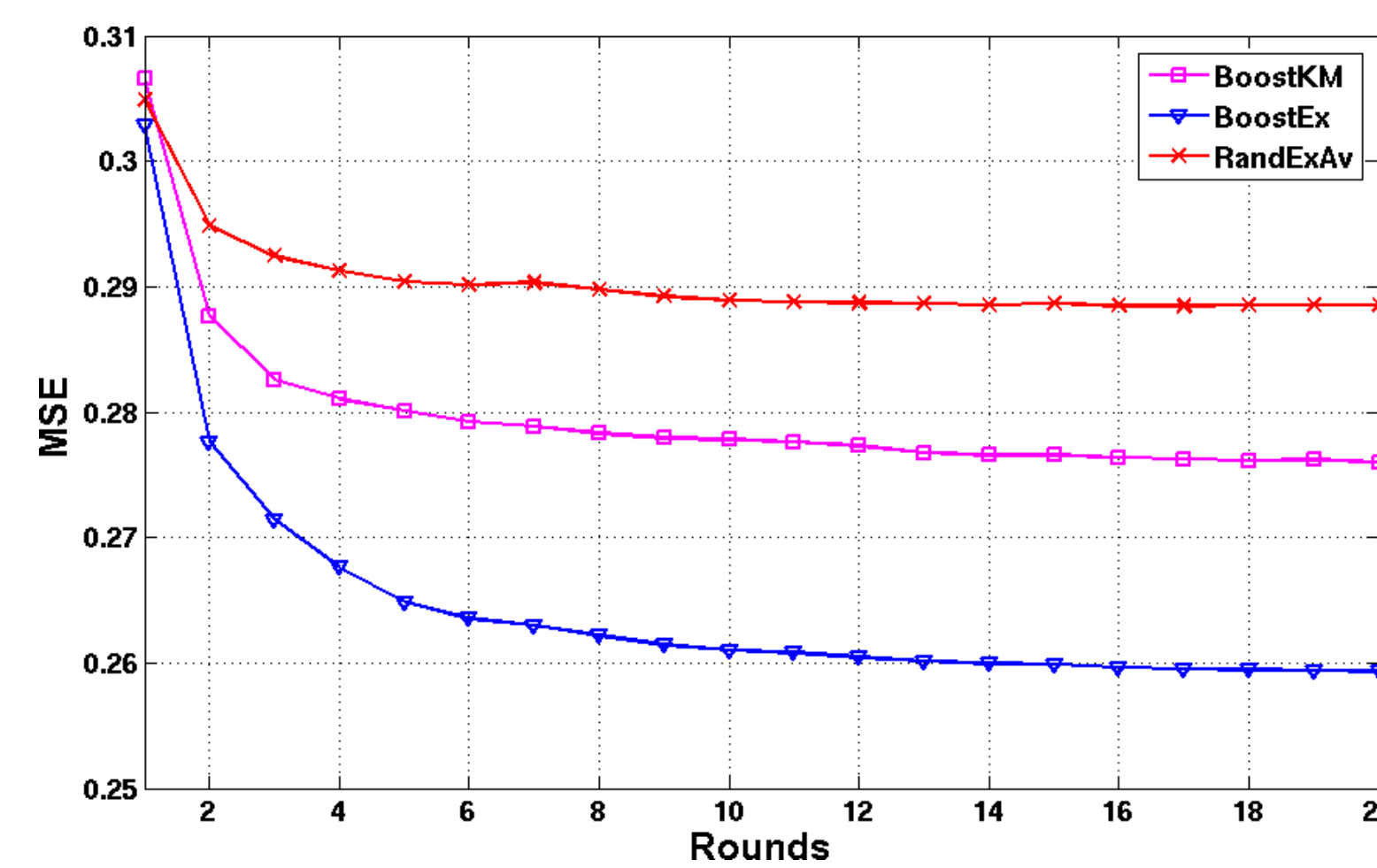
- **Advantages of Ensemble Approach[2]:**
  - **Statistical:** Better approximation of test data.
  - **Computational:** Avoid local optimal in large datasets.
  - **Representational:** Overcomes hypothesis space limitations of single model.

## Application: Image Restoration



### Degradation Aware Boosting:

- Random-projection based compressive sensing. Measurement matrices have i.i.d. Gaussian entries.
- Boosted dictionaries optimized to degradation. Knowledge of only the form of degradation is sufficient.



Error convergence of proposed ensemble models

### Weak Dictionaries Used:

- Boosted Example Selection (**BoostEx**)
  - In each round,  $D_l$  is a normalized set of  $K$  examples chosen according to  $\{p(x_i)\}$
- Boosted K-Means (**BoostKM**)

$$\min_{\{\mu_k\}_{k=1}^K, \{\mathcal{M}_k\}_{k=1}^K} \sum_{k=1}^K \sum_{i \in \mathcal{M}_k} p(x_i) \|x_i - \mu_k\|_2^2$$

- Dictionary elements are just the normalized initial cluster centers.

### Weak Representations and Boosting[3]:

- Greedy forward selection model
 
$$X_l = (1 - \alpha_l) X_{l-1} + \alpha_l D_l A_l$$
- Final approximation is an affine combination of the individual approximations.
- The weights  $\{\alpha_l\}$  have a one-to-one mapping to  $\{\beta_l\}$ .
- $D_{l+1}$  is chosen to well-represent samples poorly approximated by  $D_l$

## Compressed Recovery Of Standard Images: PSNR (Db)

Image	Number Of Measurements (N)								
	N = 8			N=16			N=32		
	Alt-Opt	KM-Boost	EX-Boost	Alt-Opt	KM-Boost	EX-Boost	Alt-Opt	KM-Boost	EX-Boost
Barbara	21.39	21.82	<b>22.05</b>	22.65	23.37	<b>23.76</b>	24.94	25.64	<b>26.34</b>
Boat	23.35	<b>23.91</b>	23.82	25.72	26.38	<b>26.42</b>	28.54	28.83	<b>29.32</b>
Couple	23.41	<b>24.04</b>	23.94	25.74	<b>26.51</b>	<b>26.51</b>	28.62	29.18	<b>29.60</b>
Fingerprint	18.40	<b>19.16</b>	18.83	21.52	<b>22.70</b>	22.34	24.97	<b>26.17</b>	26.16
House	24.84	<b>25.52</b>	25.29	27.73	<b>28.62</b>	28.44	30.97	31.66	<b>31.97</b>
Lena	25.51	<b>26.16</b>	25.94	28.12	<b>28.86</b>	28.82	30.99	31.50	<b>31.85</b>
Man	24.18	<b>24.75</b>	24.69	26.43	27.10	<b>27.23</b>	29.27	29.74	<b>30.24</b>
Peppers	21.54	<b>22.11</b>	21.99	24.08	24.60	<b>24.66</b>	27.20	27.34	<b>27.89</b>

[1] Coates, Adam, and Andrew Y. Ng. "The importance of encoding versus training with sparse coding and vector quantization." International conference on machine learning. Vol. 8. 2011.  
 [2] Dietterich, Thomas G. "Ensemble methods in machine learning." Multiple classifier systems. Springer Berlin Heidelberg, 2000. 1-15.  
 [3] Freund, Yoav, Robert Schapire, and N. Abe. "A short introduction to boosting." Journal-Japanese Society For Artificial Intelligence 14.771-780 (1999): 1612.