On Deep Multi-View Representation Learning

Weiran Wang[1] **Raman Arora**[2] **Karen Livescu**[1] **Jeff A. Bilmes**[3]

[1]TTI-Chicago [2]Johns Hopkins University [3]University of Washington

1 Overview

- **Background**
  - We consider learning representations (features) in a multi-view setting, where we have access to multiple unlabeled views for feature learning, while only one view is available for test tasks.
  - Prior theoretical and empirical results show advantages of multi-view learning, including using deep network-based approaches.
- **This work**
  - We study several old and new methods using either/both autoencoder (reconstruction error) based and canonical correlation analysis (CCA) based learning criteria.
  - We find that our newly proposed **deep canonically correlated autoencoders (DCCAE)** performs best on most tasks.
  - Code + MNIST benchmark available at `http://ttic.uchicago.edu/~wwang5/`

4 Speech and NLP experiments

- **Acoustic feature learning** from audio + articulatory measurements. Task: speaker-independent phonetic recognition, measured via phone error rate (PER, %, ↓).
- **Multilingual word embedding learning** from paired (English, German) input LSA embeddings. Task: bigram similarity (AN & VN), measured via Spearman's ρ (%, ↑).

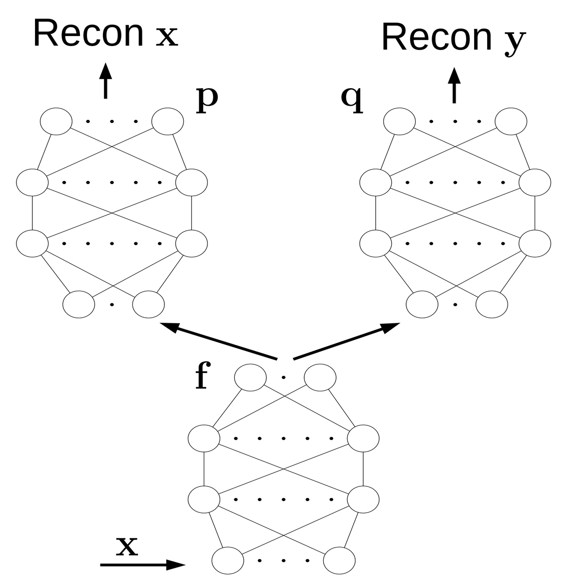
Method	PER (%)
Baseline	34.8
SplitAE	29.0
CorrAE	30.6
DistAE	33.2
CCA	26.7
KCCA	26.0
DCCA	**24.8**
DCCAE	**24.3**

Method	AN (ρ)	VN (ρ)	Avg. (ρ)
Baseline	45.0	39.1	42.1
SplitAE	47.0	**45.0**	46.0
CorrAE	43.0	42.0	42.5
DistAE	43.6	39.4	41.5
CCA	46.6	37.7	42.2
KCCA	46.4	42.9	44.7
DCCA	48.5	42.5	45.5
DCCAE	**49.1**	43.2	**46.2**

5 Conclusions

- CCA-based objectives tend to outperform autoencoders $\Longrightarrow$ no need to reconstruct inputs faithfully. Best overall is DCCAE.
- Uncorrelatedness constraint is important.
- Future: Consider stronger constraints than uncorrelatedness.

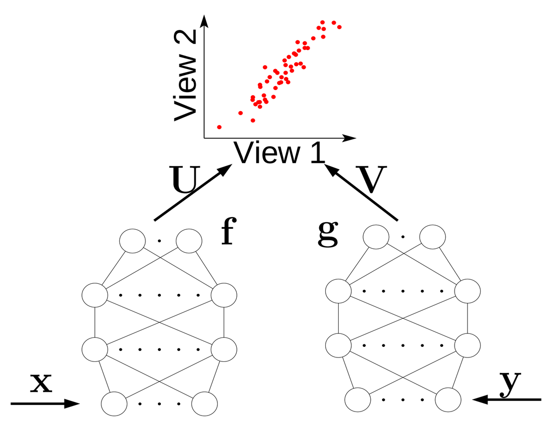
2 DNN-based multi-view feature learning



Split autoencoders (SplitAE)

[Ngiam *et al.*, 2011]

$$\min_{\mathbf{W_f},\mathbf{W_p},\mathbf{W_q}} \frac{1}{N}\sum_{i=1}^{N}\left(\|\mathbf{x}_i-\mathbf{p}(\mathbf{f}(\mathbf{x}_i))\|^2+\|\mathbf{y}_i-\mathbf{q}(\mathbf{f}(\mathbf{x}_i))\|^2\right)$$



Deep canonical correlation analysis (DCCA)

[Andrew *et al.*, 2013]

$$\max_{\mathbf{W_f},\mathbf{W_g},\mathbf{U},\mathbf{V}} \frac{1}{N}\operatorname{tr}\left(\mathbf{U}^\top\mathbf{f}(\mathbf{X})\mathbf{g}(\mathbf{Y})^\top\mathbf{V}\right)$$

$$\text{s.t.}\quad \mathbf{U}^\top\left(\frac{1}{N}\mathbf{f}(\mathbf{X})\mathbf{f}(\mathbf{X})^\top+r_x\mathbf{I}\right)\mathbf{U}=\mathbf{I},$$

$$\mathbf{V}^\top\left(\frac{1}{N}\mathbf{g}(\mathbf{Y})\mathbf{g}(\mathbf{Y})^\top+r_y\mathbf{I}\right)\mathbf{V}=\mathbf{I},$$

$$\mathbf{u}_i^\top\mathbf{f}(\mathbf{X})\mathbf{g}(\mathbf{Y})^\top\mathbf{v}_j=0,\quad\text{for}\quad i\neq j,$$

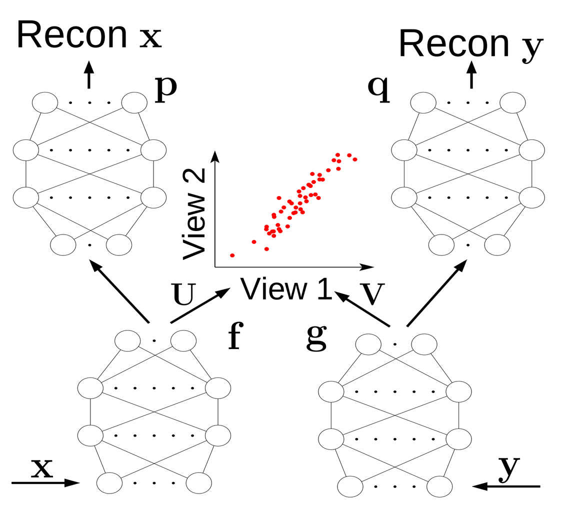
Deep canonically correlated autoencoders (DCCAE)

$$\min_{\mathbf{W_f},\mathbf{W_g},\mathbf{W_p},\mathbf{W_q},\mathbf{U},\mathbf{V}} -\frac{1}{N}\operatorname{tr}\left(\mathbf{U}^\top\mathbf{f}(\mathbf{X})\mathbf{g}(\mathbf{Y})^\top\mathbf{V}\right)$$

$$+\frac{\lambda}{N}\sum_{i=1}^{N}\left(\|\mathbf{x}_i-\mathbf{p}(\mathbf{f}(\mathbf{x}_i))\|^2+\|\mathbf{y}_i-\mathbf{q}(\mathbf{g}(\mathbf{y}_i))\|^2\right)$$

s.t. the same constraints in DCCA.

- Stochastic optimization can be applied to DCCA/DCCAE.



Correlated autoencoders (CorrAE)

$$\min_{\mathbf{W_f},\mathbf{W_g},\mathbf{W_p},\mathbf{W_q},\mathbf{U},\mathbf{V}} -\frac{1}{N}\operatorname{tr}\left(\mathbf{U}^\top\mathbf{f}(\mathbf{X})\mathbf{g}(\mathbf{Y})^\top\mathbf{V}\right)$$

$$+\frac{\lambda}{N}\sum_{i=1}^{N}\left(\|\mathbf{x}_i-\mathbf{p}(\mathbf{f}(\mathbf{x}_i))\|^2+\|\mathbf{y}_i-\mathbf{q}(\mathbf{g}(\mathbf{y}_i))\|^2\right)$$

$$\text{s.t.}\quad \mathbf{u}_i^\top\mathbf{f}(\mathbf{X})\mathbf{f}(\mathbf{X})^\top\mathbf{u}_i=\mathbf{v}_i^\top\mathbf{g}(\mathbf{Y})\mathbf{g}(\mathbf{Y})^\top\mathbf{v}_i=N,$$

- CorrAE constraint is a relaxed version of that of DCCAE.

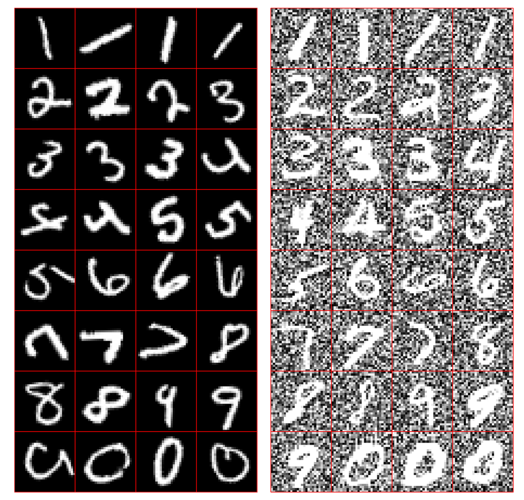
Minimum-distance autoencoders (DistAE)

$$\min_{\mathbf{W_f},\mathbf{W_g},\mathbf{W_p},\mathbf{W_q}} \frac{1}{N}\sum_{i=1}^{N}\frac{\|\mathbf{f}(\mathbf{x}_i)-\mathbf{g}(\mathbf{y}_i)\|^2}{\|\mathbf{f}(\mathbf{x}_i)\|^2+\|\mathbf{g}(\mathbf{y}_i)\|^2}$$

$$+\frac{\lambda}{N}\sum_{i=1}^{N}\left(\|\mathbf{x}_i-\mathbf{p}(\mathbf{f}(\mathbf{x}_i))\|^2+\|\mathbf{y}_i-\mathbf{q}(\mathbf{g}(\mathbf{y}_i))\|^2\right)$$

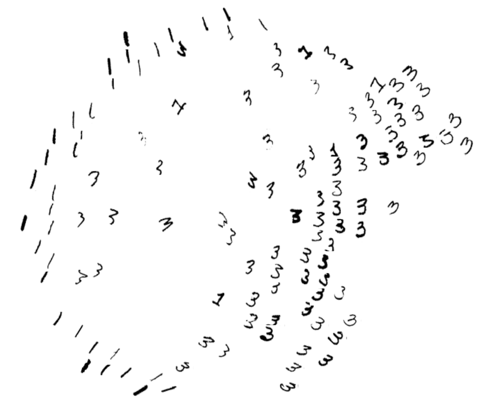
- Objective is unconstrained and decouples over samples.

3 Noisy MNIST experiments

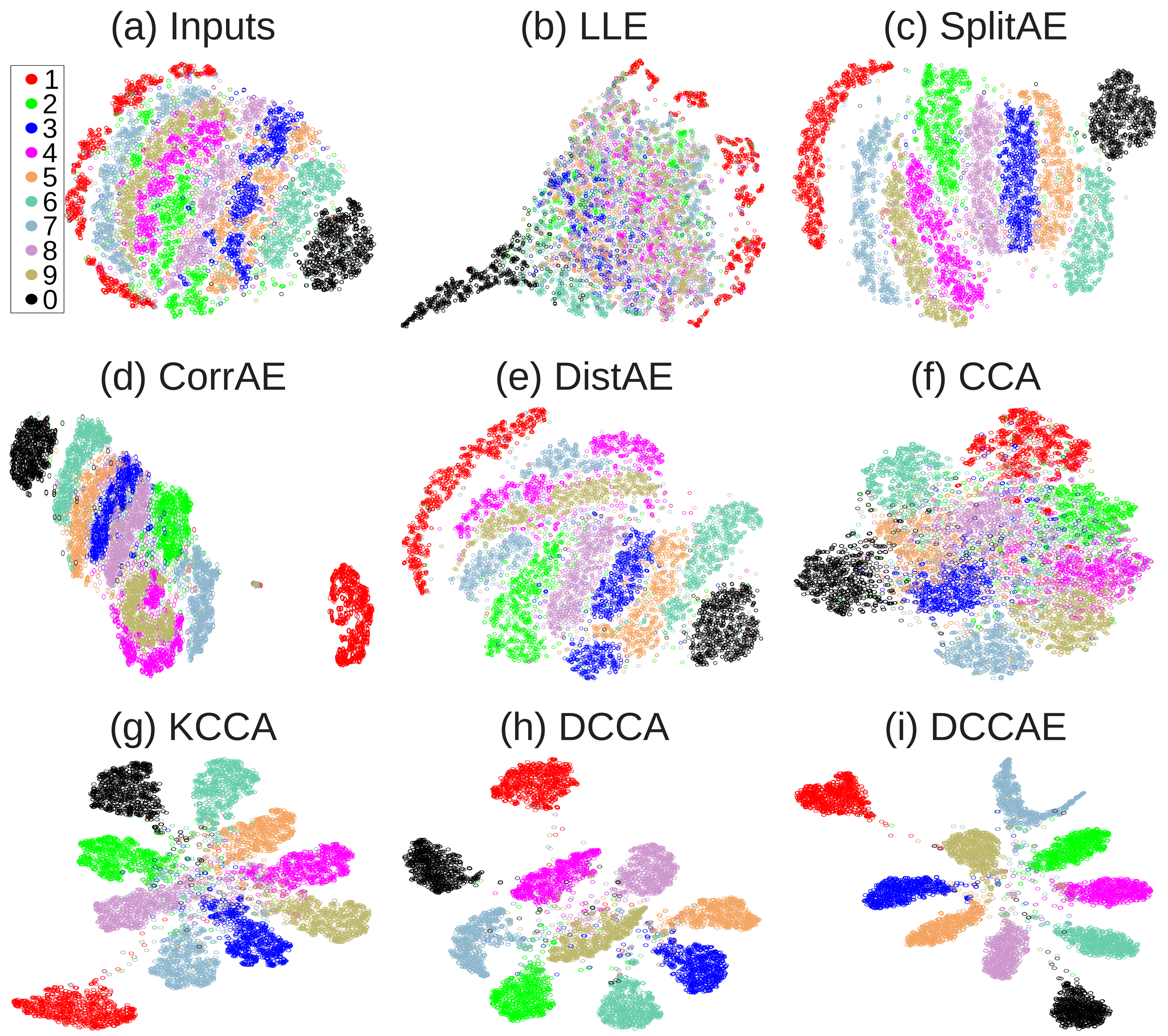


View 1: randomly rotated digits, $\theta\in[-\pi/4,\pi/4]$. View 2: randomly chosen image of same identity + pixel noise. **The views are uncorrelated given the label.**

View 1 t-SNE



- We tune the learned projections to maximize clustering accuracy, and then learn a linear SVM on the best projection.



t-SNE visualizations of learned View 1 projections (~ 10D).

- CCAs better separate different classes while suppressing the rotational variation (useless for discrimination).
- Uncorrelatedness constraint in CCA is essential (CorrAE vs. DCCAE).
- CCA objective outperforms reconstruction objective.

Method	SVM Error (%)
Baseline	13.1
SplitAE	11.9
CorrAE	12.9
DistAE	16.0
CCA	19.6
KCCA	4.5
DCCA	**2.9**
DCCAE	**2.2**