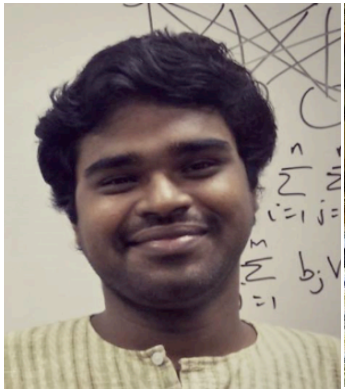


Finding Structure in Data

(Bridge) Correlational Neural Networks

B. RAVINDRAN

RECONFIGURABLE AND INTELLIGENT SYSTEMS ENGINEERING (**RISE**) GROUP

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AND INTERDISCIPLINARY LABORATORY FOR DATA SCIENCES (**ILDS**)

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

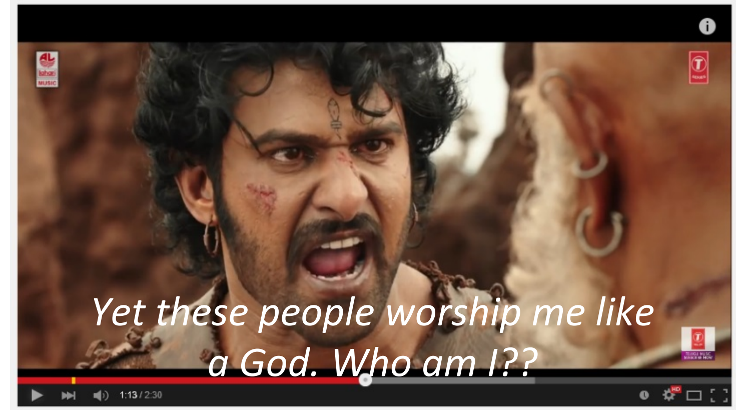
JOINT WORK WITH

SARATH CHANDAR A.P. UNIVERSITE DE MONTREAL

JANARTHANAN R., IIT MADRAS

MITESH KHAPRA IBM IRL /IIT MADRAS

HUGO LAROCHELLE TWITTER CORTEX, USA., UNIVERSITE DE SHERBROOKE

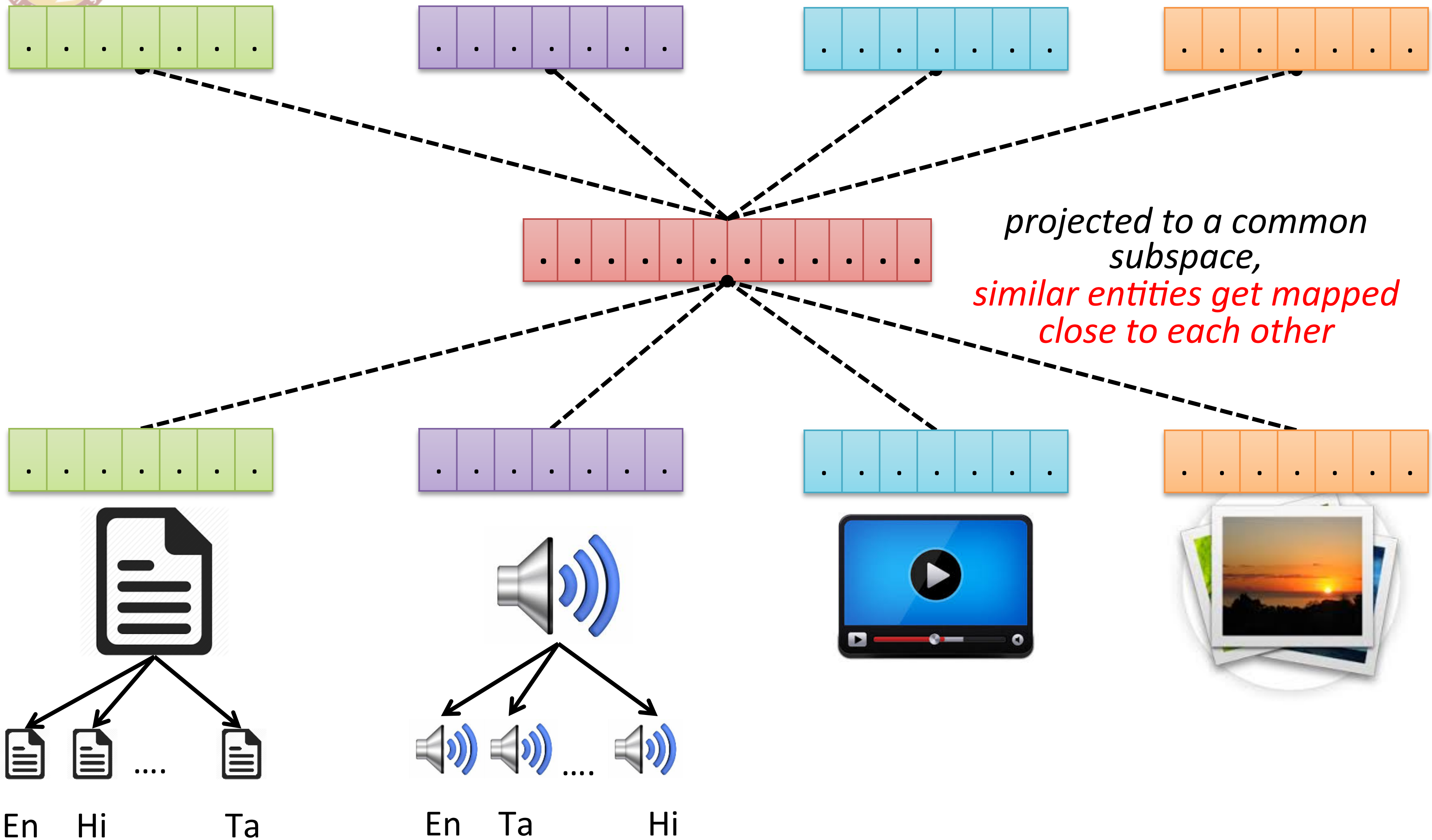


Video, Tamil Audio, English subtitles

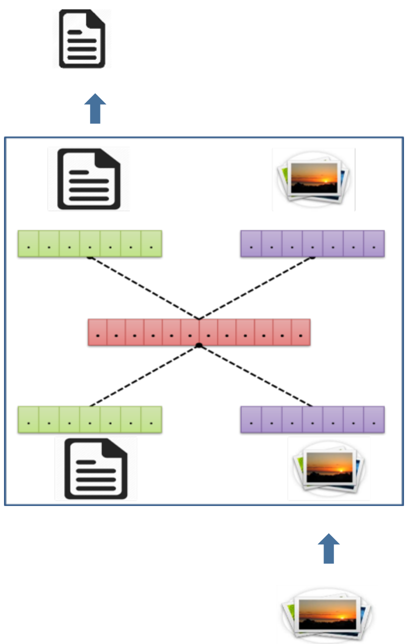
English: brown horses eating tall grass beside a body of water

French: chevaux brun manger l'herbe haute à côté d'un corps de l'eau

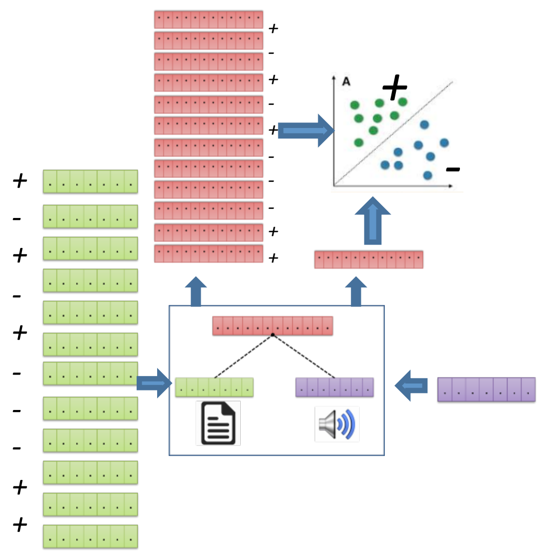
We live in an increasingly multilingual multimodal world reconstruct views



Learning common representations for multiple views

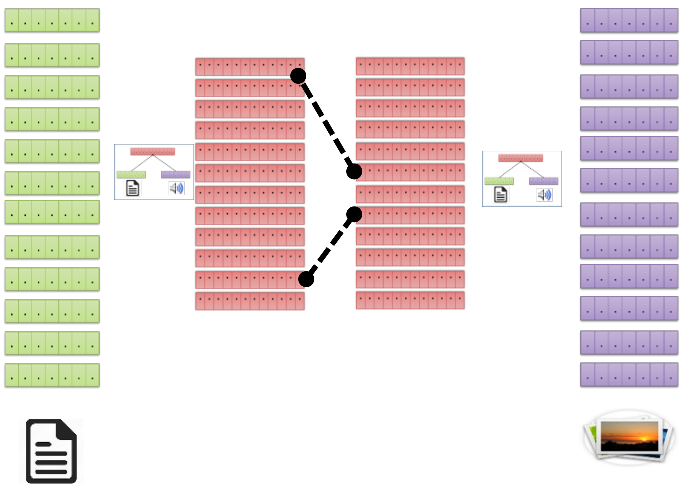


(i) reconstruct a missing view



Profanity labeled text

(ii) Transfer learning between views



(iii) Cross view retrieval/matching

Why learn such common representations?

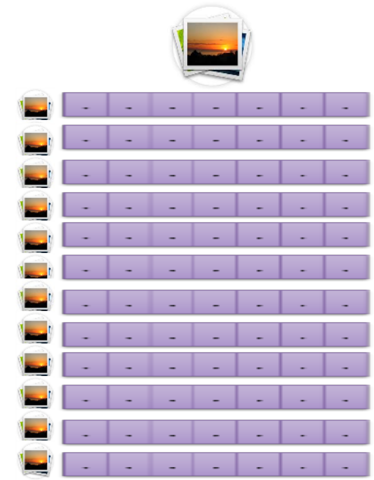
Outline

- Related Work
- Proposed Model: CorrNet
- Analysis of the model
- Application to cross language learning
- Beyond 2 views – Bridge CorrNet
- Results

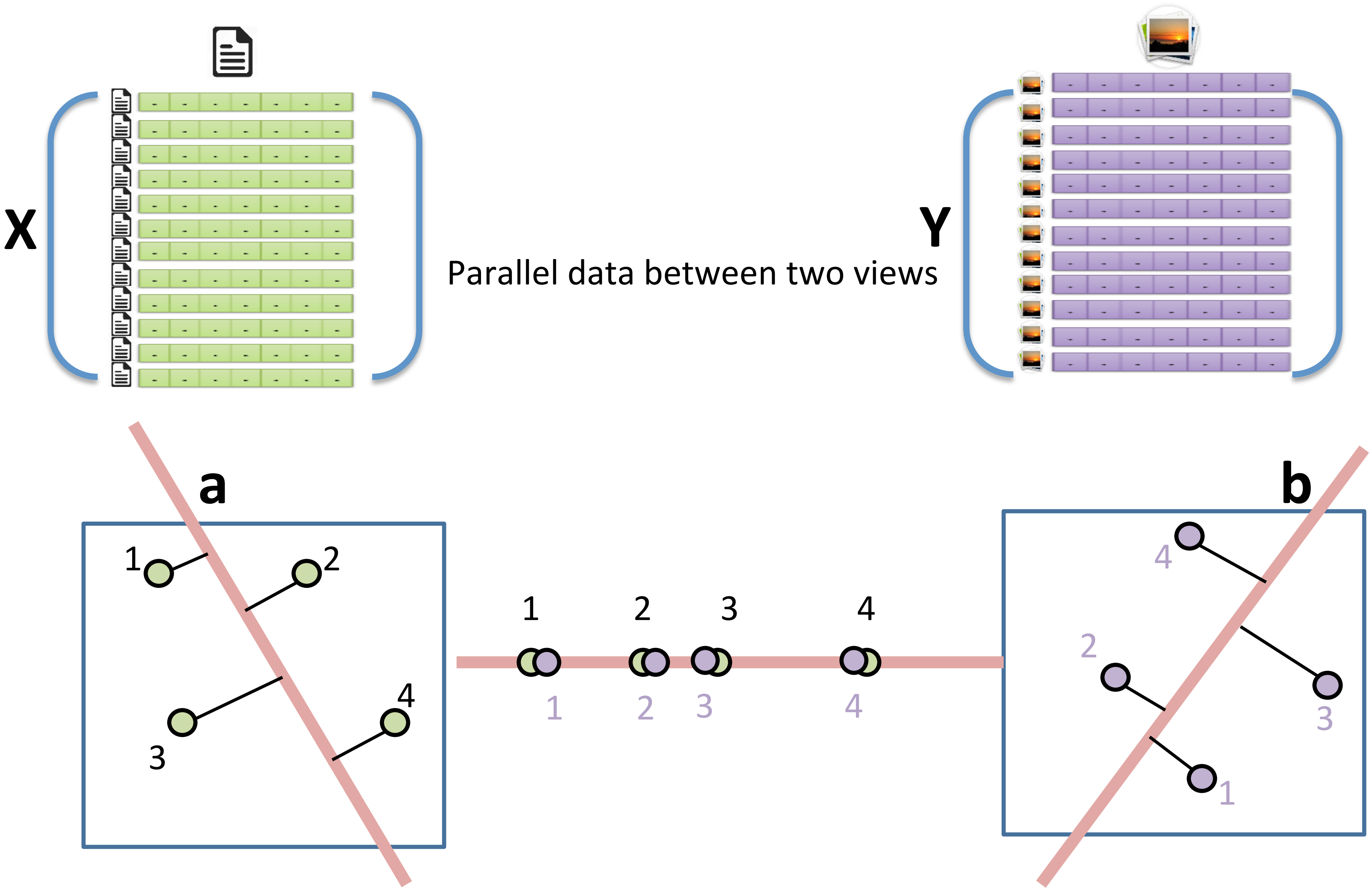
Outline

- **Related Work**
- Proposed Model: CorrNet
- Analysis of the model
- Application to cross language learning
- Beyond 2 views – Bridge CorrNet
- Results

Parallel data between two views



Any common representation learning method would need this…



Canonical Correlation Analysis (CCA)

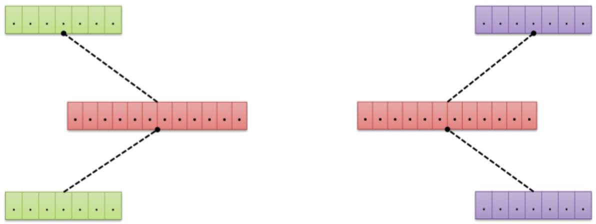
This looks good… then why go beyond CCA?

- Scalability issues
- Non-trivial to extend beyond 2 views
- Lacks reconstruction capabilities
  - *A* and *B* need not be invertible
  - Only a low rank approximate reconstruction would be possible

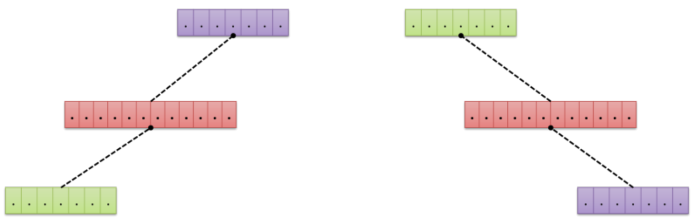
Outline

- Related Work
- **Proposed Model: CorrNet**
- Analysis of the model
- Application to cross language learning
- Beyond 2 views – Bridge CorrNet
- Results

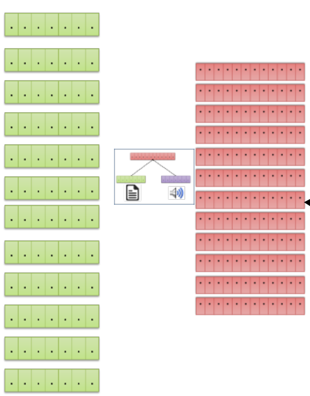
Wishlist



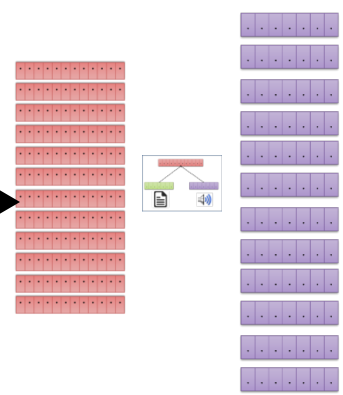
(i) good self reconstruction



(ii) good cross reconstruction

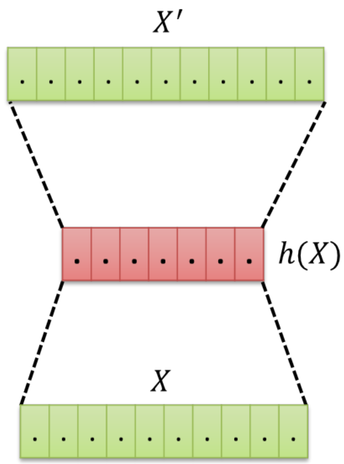


(iii)High correlation



(iv) scalable

(v) easily extendable to multiple views

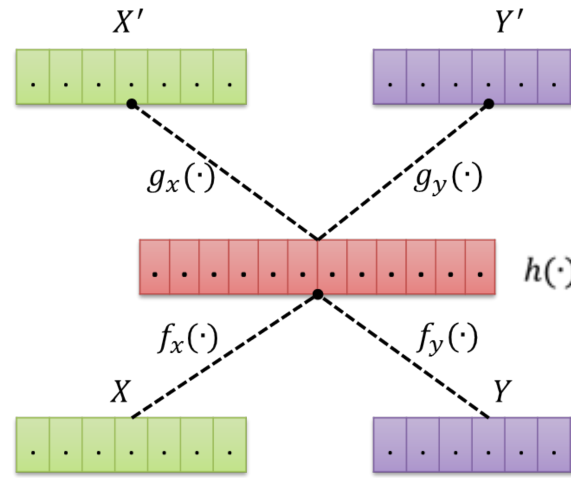


$$h(X) = f(X) = \sigma(\mathbf{W}X + b)$$ *encoder*

$$X' = g(h(X)) = \sigma(\mathbf{W}'h(X) + b')$$ *decoder*

$$\min \sum_{i=1}^{N} \left(X_i - g(h(X_i)) \right)^2$$ *use back propogation*

Background: a neural network based single view autoencoder



encoder

$$h_x(X) = f_x(X) = \sigma(\mathbf{W}_x X + b)$$

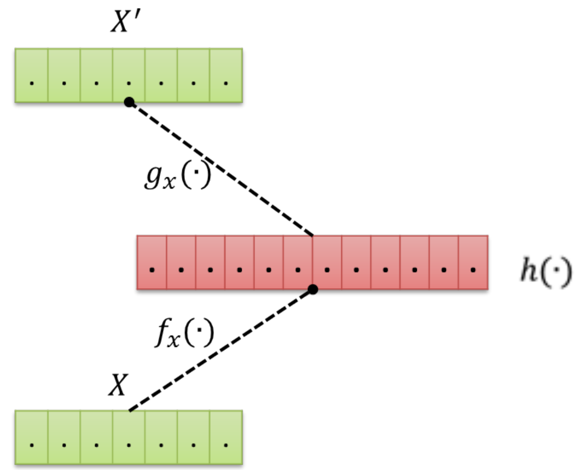
$$h_y(Y) = f_y(Y) = \sigma(\mathbf{W}_y Y + b)$$

decoder

$$X' = g_x(h_x(X)) = \sigma(\mathbf{W}_x' h_x(X) + b')$$

$$X' = g_y(h_y(Y)) = \sigma(\mathbf{W}_y' h_y(Y) + b')$$

Correlational Neural Network



A multiview autoencoder

$$\min \sum_{i=1}^{N} \left(X_i - g_x(f_x(X_i)) \right)^2$$

encoder

$$h_x(X) = f_x(X) = \sigma(\mathbf{W}_x X + b)$$

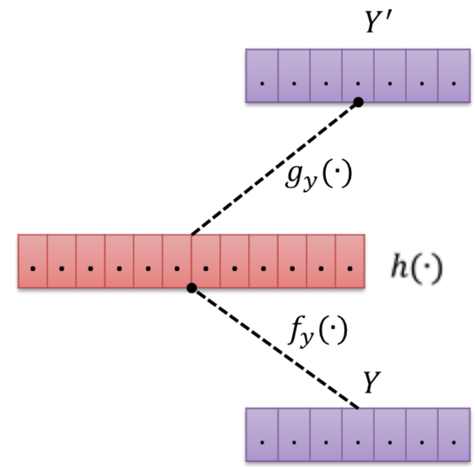
$$h_y(Y) = f_y(Y) = \sigma(\mathbf{W}_y Y + b)$$

decoder

$$X' = g_x(h_x(X)) = \sigma(\mathbf{W}'_x h_x(X) + b')$$

$$X' = g_y(h_y(Y)) = \sigma(\mathbf{W}'_y h_y(Y) + b')$$

Correlational Neural Network



A multiview autoencoder

encoder

$$h_x(X) = f_x(X) = \sigma(\mathbf{W}_x X + b)$$

$$h_y(Y) = f_y(Y) = \sigma(\mathbf{W}_y Y + b)$$

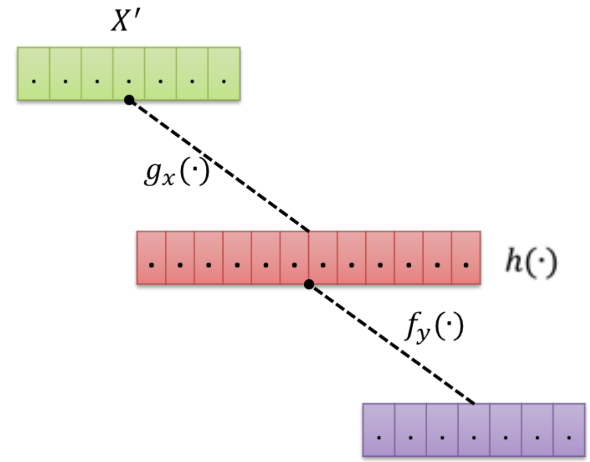
decoder

$$X' = g_x(h_x(X)) = \sigma(\mathbf{W}'_x h_x(X) + b')$$

$$X' = g_y(h_y(Y)) = \sigma(\mathbf{W}'_y h_y(Y) + b')$$

$$\min \sum_{i=1}^{N} \left(X_i - g_x(f_x(X_i))\right)^2 + \sum_{i=1}^{N} \left(Y_i - g_y(f_y(Y_i))\right)^2$$

Correlational Neural Network



A multiview autoencoder

encoder

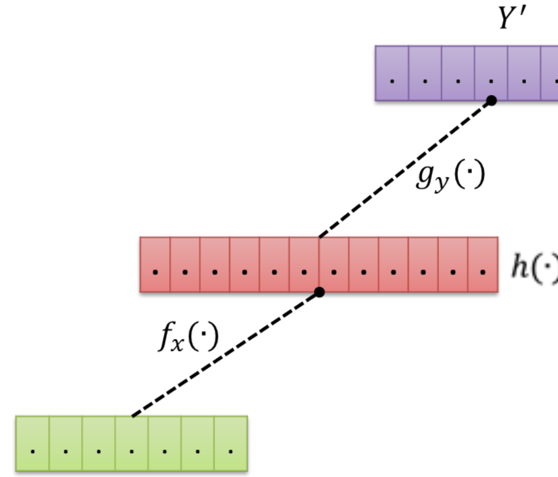
$$h_x(X) = f_x(X) = \sigma(\mathbf{W}_x X + b)$$
$$h_y(Y) = f_y(Y) = \sigma(\mathbf{W}_y Y + b)$$

decoder

$$X' = g_x(h_x(X)) = \sigma(\mathbf{W}_x' h_x(X) + b')$$
$$X' = g_y(h_y(Y)) = \sigma(\mathbf{W}_y' h_y(Y) + b')$$

$$\min \sum_{i=1}^{N} \left(X_i - g_x(f_x(X_i))\right)^2 + \sum_{i=1}^{N} \left(Y_i - g_y(f_y(Y_i))\right)^2 + \sum_{i=1}^{N} \left(X_i - g_x(f_y(Y_i))\right)^2$$

Correlational Neural Network



A multiview autoencoder

encoder

$$h_x(X) = f_x(X) = \sigma(\mathbf{W}_x X + b)$$

$$h_y(Y) = f_y(Y) = \sigma(\mathbf{W}_y Y + b)$$

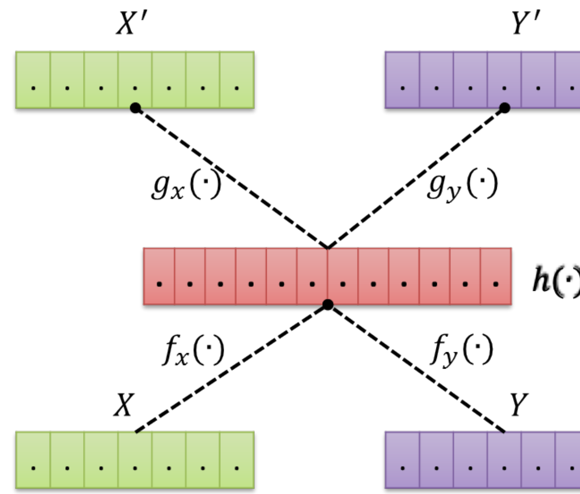
decoder

$$X' = g_x(h_x(X)) = \sigma(\mathbf{W}'_x h_x(X) + b')$$

$$X' = g_y(h_y(Y)) = \sigma(\mathbf{W}'_y h_y(Y) + b')$$

$$\min \sum_{i=1}^{N} \left(X_i - g_x(f_x(X_i))\right)^2 + \sum_{i=1}^{N} \left(Y_i - g_y(f_y(Y_i))\right)^2 + \sum_{i=1}^{N} \left(X_i - g_x(f_y(Y_i))\right)^2 + \sum_{i=1}^{N} \left(Y_i - g_y(f_x(X_i))\right)^2$$

Correlational Neural Network



A multiview autoencoder

encoder

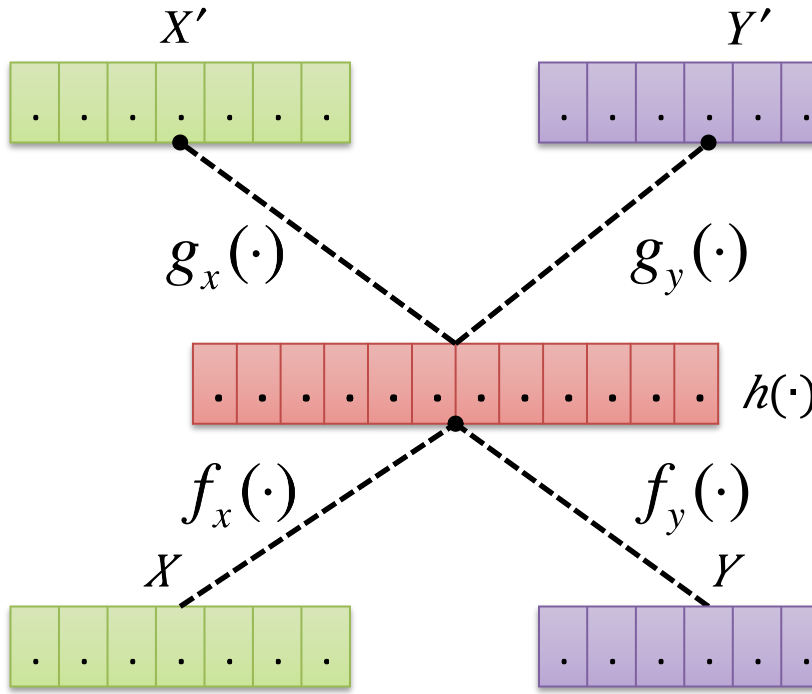
$$h_x(X) = f_x(X) = \sigma(\mathbf{W}_x X + b)$$
$$h_y(Y) = f_y(Y) = \sigma(\mathbf{W}_y Y + b)$$

decoder

$$X' = g_x(h_x(X)) = \sigma(\mathbf{W}'_x h_x(X) + b')$$
$$X' = g_y(h_y(Y)) = \sigma(\mathbf{W}'_y h_y(Y) + b')$$

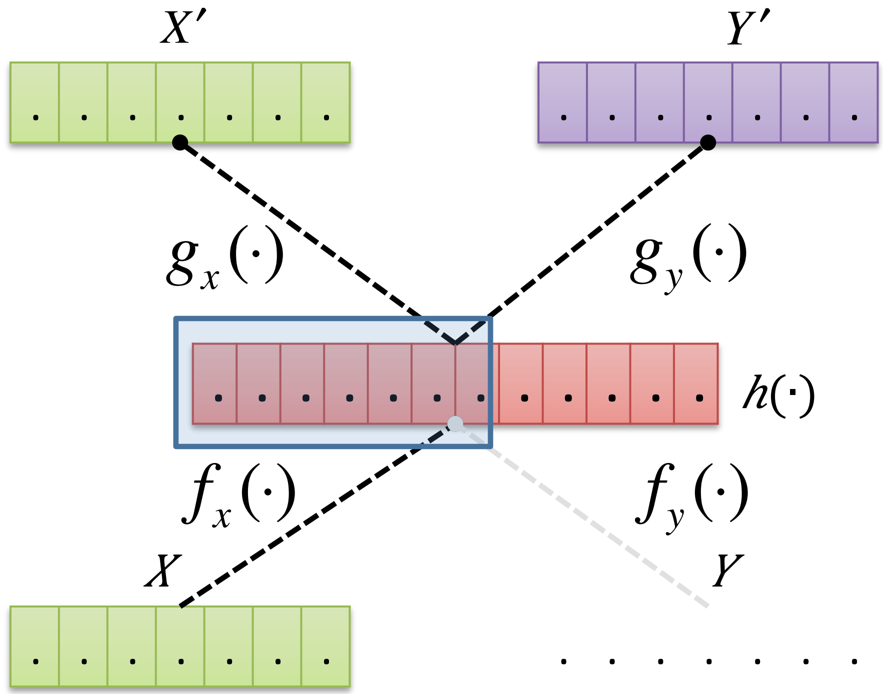
$$\min \sum_{i=1}^{N}\left(X_i - g_x(f_x(X_i))\right)^2 + \sum_{i=1}^{N}\left(Y_i - g_y(f_y(Y_i))\right)^2 + \sum_{i=1}^{N}\left(X_i - g_x(f_y(Y_i))\right)^2 + \sum_{i=1}^{N}\left(Y_i - g_y(f_x(X_i))\right)^2$$

Correlational Neural Network



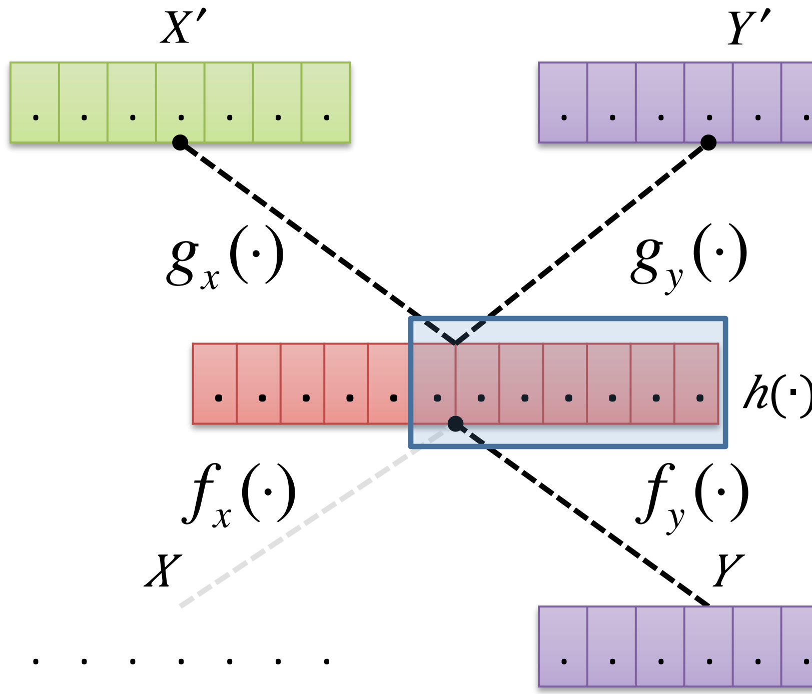
So far so good…. But will the representations h(X) and h(Y) be correlated?

Turns out that there is no guarantee for this !



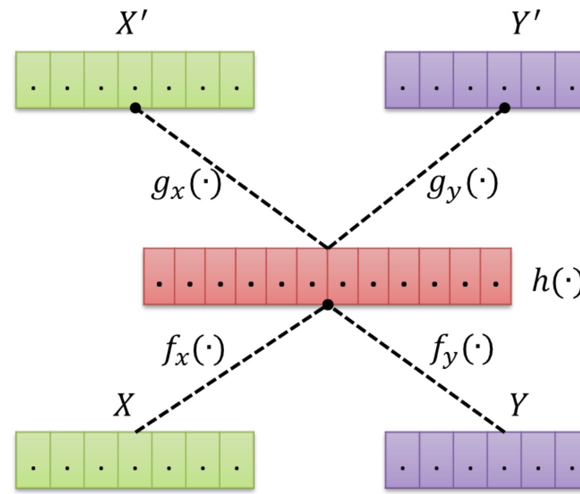
So far so good…. But will the representations h(X) and h(Y) be correlated?

Turns out that there is no guarantee for this !



So far so good.... But will the representations h(X) and h(Y) be correlated?

Turns out that there is no guarantee for this !



A multiview autoencoder

encoder

$$h_x(X) = f_x(X) = \sigma(\mathbf{W}_x X + b)$$

$$h_y(Y) = f_y(Y) = \sigma(\mathbf{W}_y Y + b)$$

decoder

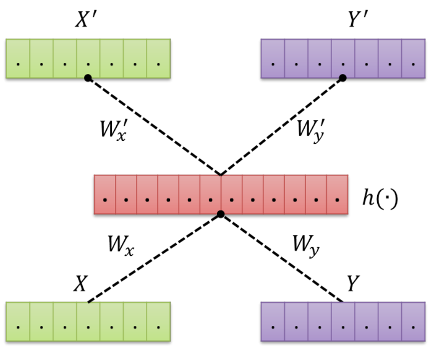
$$X' = g_x(h_x(X)) = \sigma(\mathbf{W}'_x h_x(X) + b')$$

$$X' = g_y(h_y(Y)) = \sigma(\mathbf{W}'_y h_y(Y) + b')$$

$$\min \sum_{i=1}^{N} \left(X_i - g_x(f_x(X_i))\right)^2 + \sum_{i=1}^{N} \left(Y_i - g_y(f_y(Y_i))\right)^2 + \sum_{i=1}^{N} \left(X_i - g_x(f_y(Y_i))\right)^2 + \sum_{i=1}^{N} \left(Y_i - g_y(f_x(X_i))\right)^2$$

$$\boxed{-corr\left(h(X), h(Y)\right)}$$

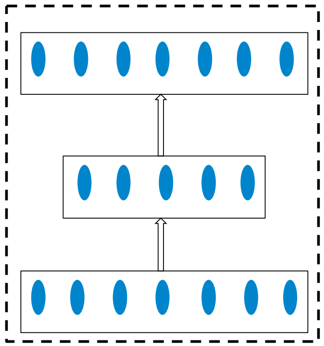
Correlational Neural Network



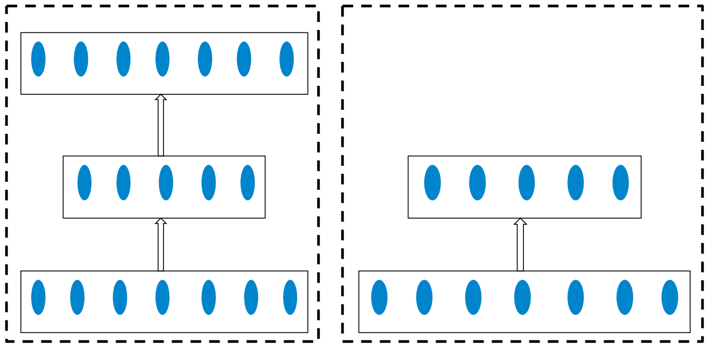
Step 1: Train a shallow Cornet

Building Deep Correlational Neural Networks

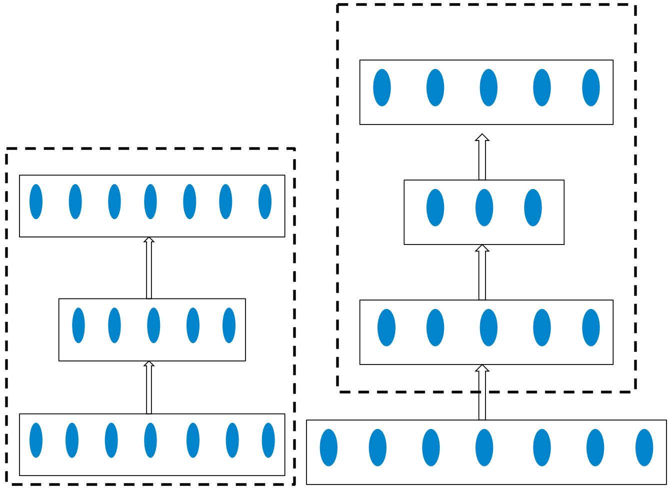
Greedy Unsupervised Layerwise pretraining (Hinton et al., 2006; Bengio et al., 2007)



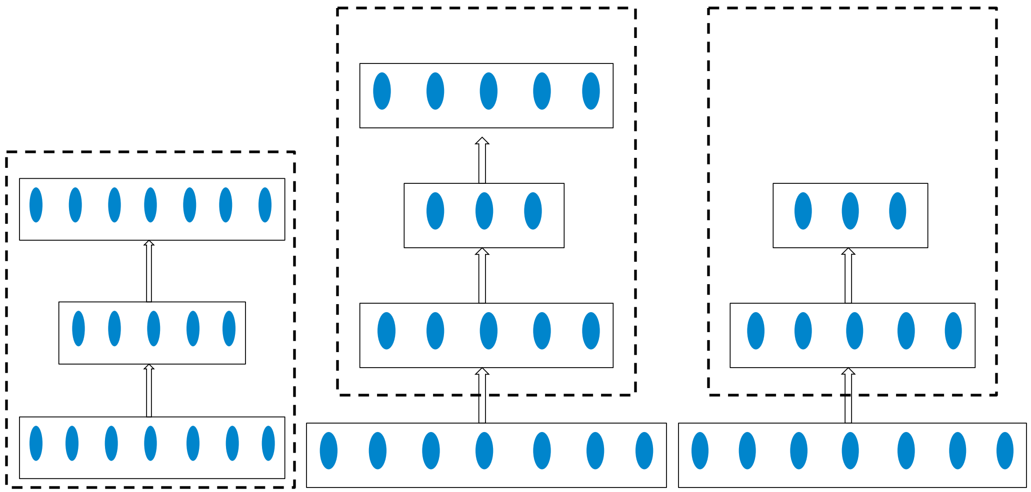
Greedy Unsupervised Layerwise pretraining
(Hinton et al., 2006; Bengio et al., 2007)



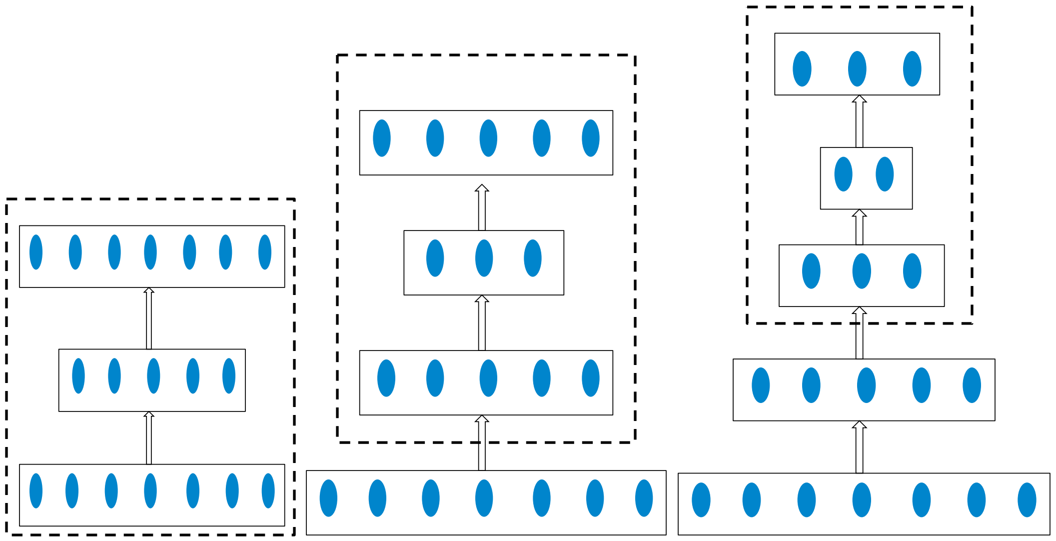
Greedy Unsupervised Layerwise pretraining
(Hinton et al., 2006; Bengio et al., 2007)



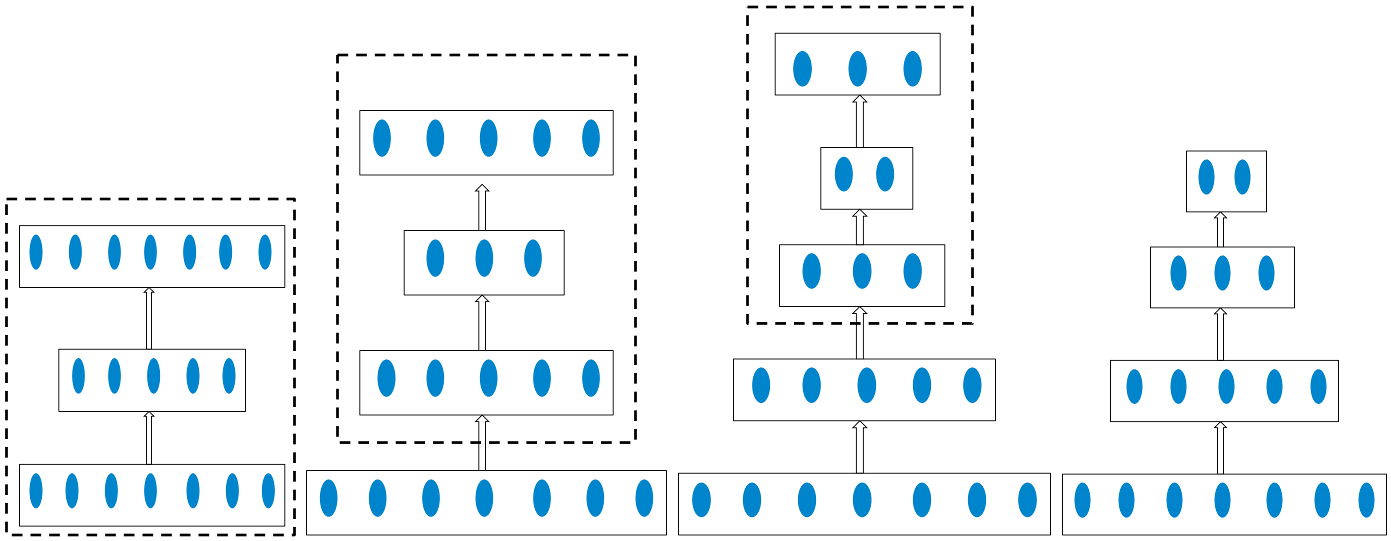
Greedy Unsupervised Layerwise pretraining
(Hinton et al., 2006; Bengio et al., 2007)



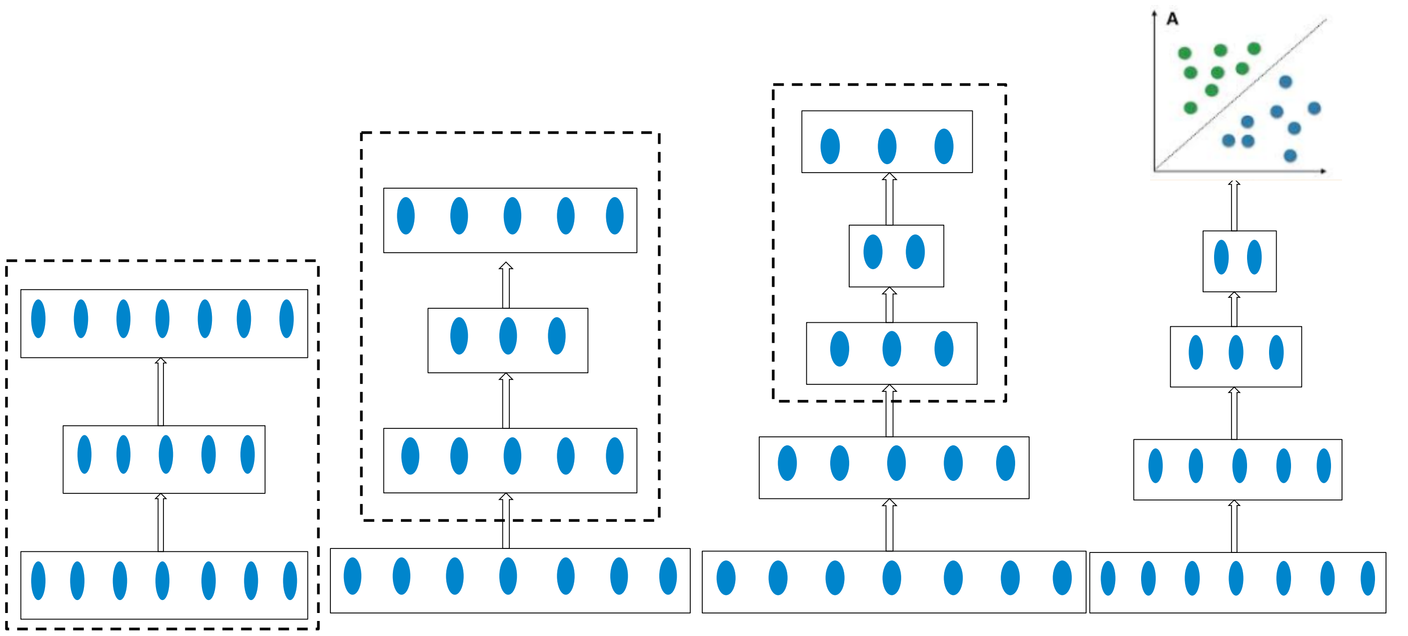
Greedy Unsupervised Layerwise pretraining
(Hinton et al., 2006; Bengio et al., 2007)



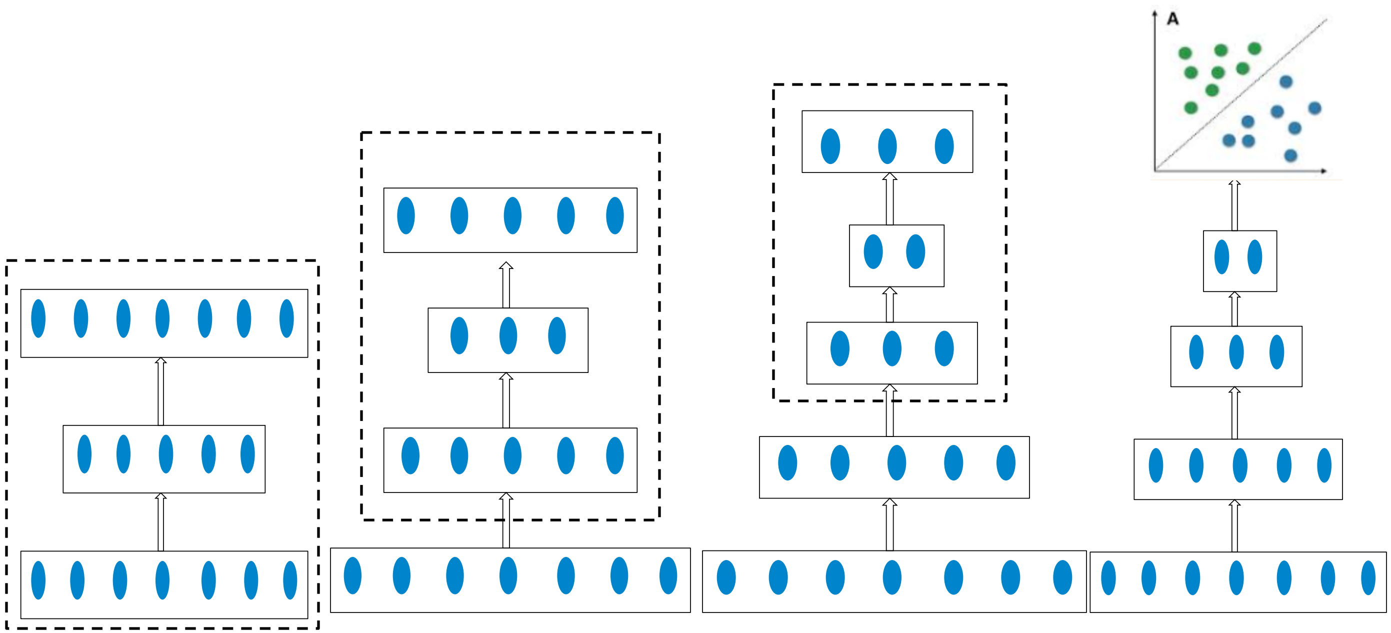
Greedy Unsupervised Layerwise pretraining
(Hinton et al., 2006; Bengio et al., 2007)



Greedy Unsupervised Layerwise pretraining
(Hinton et al., 2006; Bengio et al., 2007)

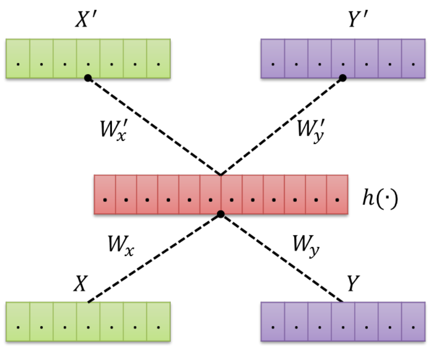


Greedy Unsupervised Layerwise pretraining
(Hinton et al., 2006; Bengio et al., 2007)

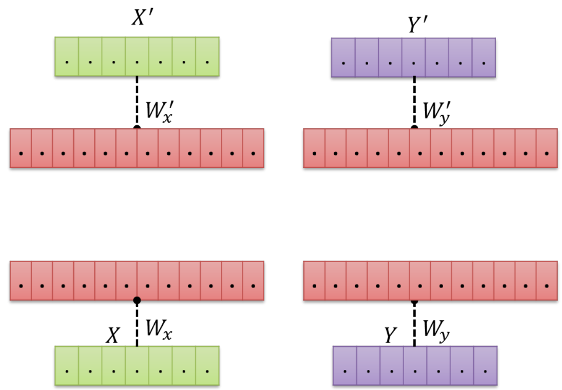


Greedy layerwise pretraining

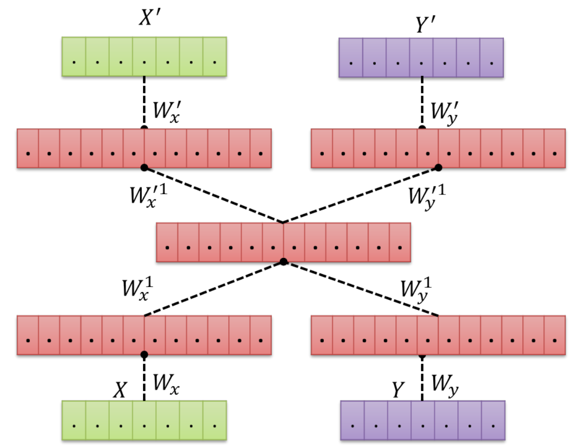
Fine-tuning



Step 1: Train a shallow Cornet



Step 2: Decompose the common hidden layer but copy the weights learned earlier



Step 3: Connect a new hidden layer and re-train

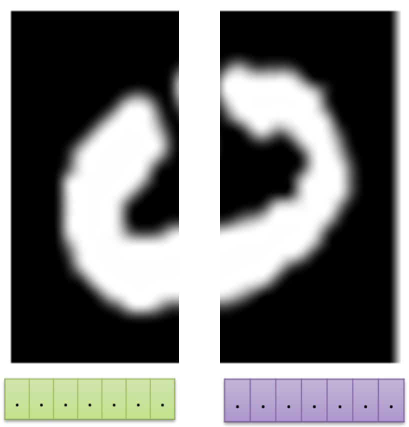
Building Deep Correlational Neural Networks

Outline

- Related Work
- Proposed Model: CorrNet
- **Analysis of the model**
- Application to cross language learning
- Beyond 2 views – Bridge CorrNet
- Results

Task

MNIST Digit Recognition



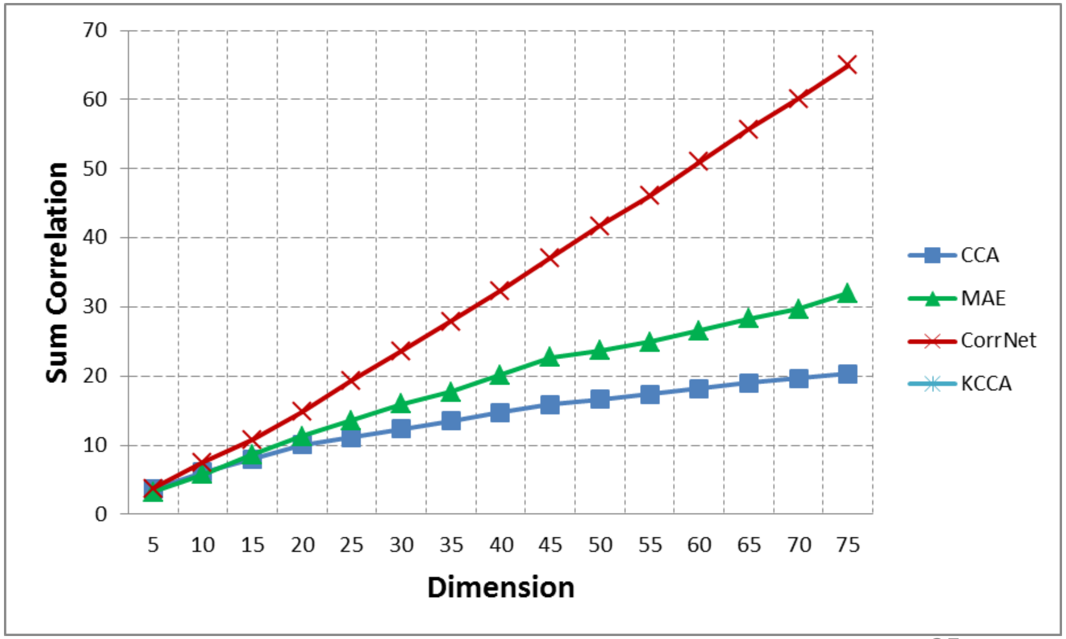
Left and right half of the image are two views of the same data

- Data
  - 60K train
  - 10K test
  - 28*28 pixels
  - R: 14*28 dim
  - L: 14*28 dim
- Evaluate
  - ability to learn correlated representations
  - effectiveness in transfer learning
  - ability to do self and cross reconstruction
  - whether adding layers helps

Ability to learn correlated representations

Model	Sum Correlation
CCA	17.05
KCCA	30.58
MAE	24.40
CorrNet	**45.47**

- *Learn 50 dimensional representations*
- *Calculate sum of correlation across all dimensions*



Effectiveness in transfer learning

Model	Left to Right	Right to Left
CCA	66.13	66.71
KCCA	70.68	70.83
MAE	68.69	72.54
CorrNet	**76.60**	**79.51**
Self learning	81.62	80.06

- *Train a digit classifier using common representation of only right half of train images*
- *Test the classifier using common representation of only left half of test images*
- *and vice versa*

Ability to do self and cross reconstruction

Model	MSE for self reconstruction	MSE for cross reconstruction
CorrNet	3.6	4.3
MAE	**2.1**	**4.2**

- *CorrNet's ability to self reconstruct is sacrificed in the interest of learning correlated representations*

Improvements using Deep CorrNet

Model	Left to Right	Right to Left	Sum Correlation
CCA	66.13	66.71	17.05
KCCA	70.68	70.83	30.58
MAE	68.69	72.54	24.40
CorrNet-50	76.60	79.51	45.47
CorrNet-500-50	77.68	77.95	**47.21**
DCCA-500-50	66.41	64.65	33.00
CorrNet-500-300-50	**80.46**	**80.47**	45.63
DCCA-500-300-50	70.06	72.43	33.77

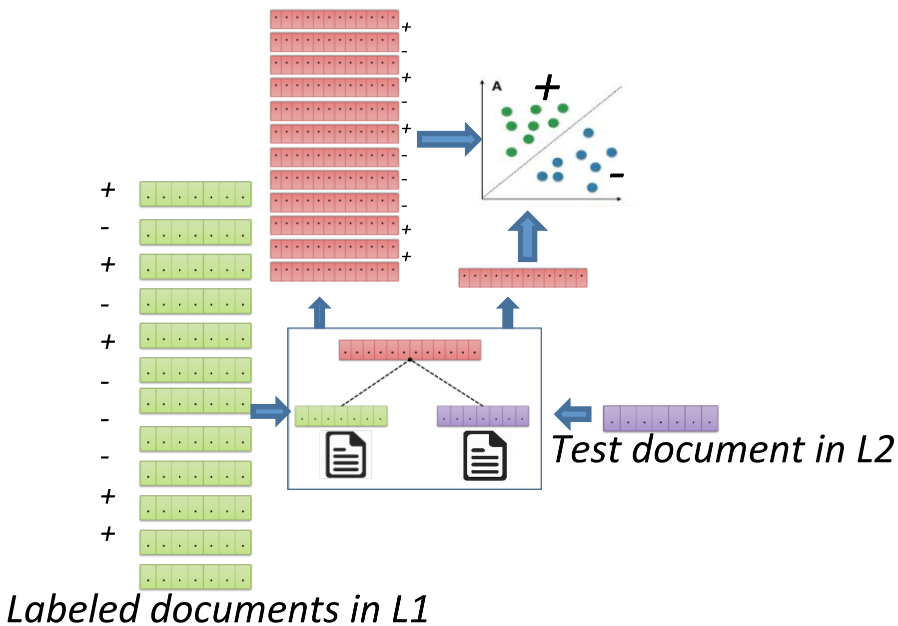
- *The Deeper versions perform better than the shallow CorrNet*

Outline

- Related Work
- Proposed Model: CorrNet
- Analysis of the model
- **Application to cross language learning**
- Beyond 2 views – Bridge CorrNet
- Results

Cross Language Learning

- Train a document classifier using labeled data in language L1
- Use this classifier to classify documents in L2



Setup

- **Language Pairs:** English<-->French, English<-->German, English<-->Spanish
- **Parallel Data:** Europarl corpus used for learning common representations
- **Labeled Data:** Reuters RCV1/RCV2 corpus with 4 classes (CCAT, ECAT, GCAT, MCAT)
- **Hyperparameteres:** dim=40, 20 epochs, stochastic gradient descent with mini-batch size 20

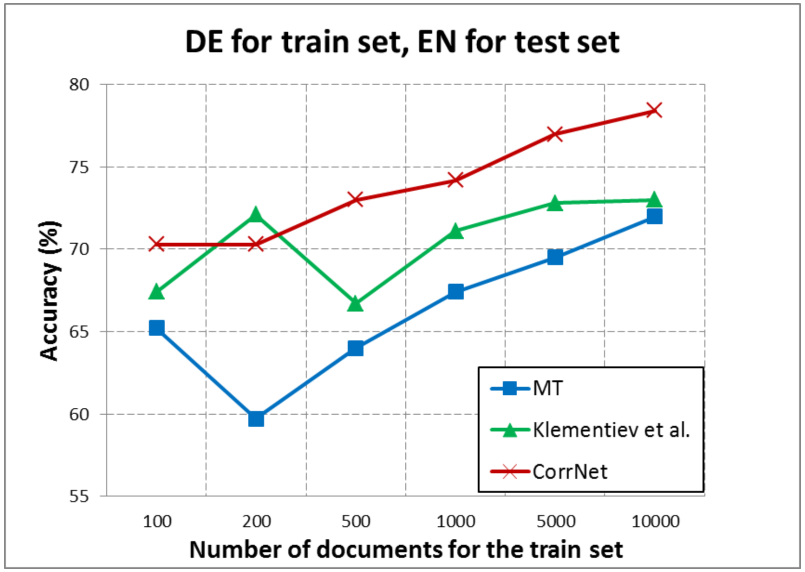
Nearest cross language neighbors

january		president		said	
en	de	en	de	en	de
january	januar	president	präsident	said	gesagt
march	märz	i	präsidentin	told	sagte
october	oktober	mr	präsidenten	say	sehr
july	juli	presidents	herr	believe	heute
december	dezember	thank	ich	saying	sagen
1999	jahres	president-in-office	ratspräsident	wish	heutigen
june	juni	report	danken	shall	letzte
month	1999	voted	danke	again	hier
oil		**microsoft**		**market**	
en	de	en	de	en	de
oil	öl	microsoft	microsoft	market	markt
supply	boden	cds	cds	markets	marktes
supplies	befindet	insider	warner	single	märkte
gas	gerät	ibm	tageszeitungen	commercial	binnenmarkt
fuel	erdöl	acquisitions	ibm	competition	märkten
mineral	infolge	shareholding	handelskammer	competitive	handel
petroleum	abhängig	warner	exchange	business	öffnung
crude	folge	online	veranstalter	goods	binnenmarktes

- *Nearest neighbors are either translations or semantically related words*

Results

Model	En→De	De→En	En→Fr	Fr→En	En→Es	Es→En
CorrNet	**91.8**	74.2	**84.6**	**74.2**	49.0	**64.4**
Machine Translation	68.1	67.4	76.3	71.1	**52.0**	58.4
Klementiev et.al. 2012	77.6	71.1	74.5	61.9	31.3	63.0
Hermann and Blunsom, 2014	88.1	**79.1**	N.A.	N.A.	N.A.	N.A.
Majority Class	46.8	46.8	22.5	25.0	15.3	22.2



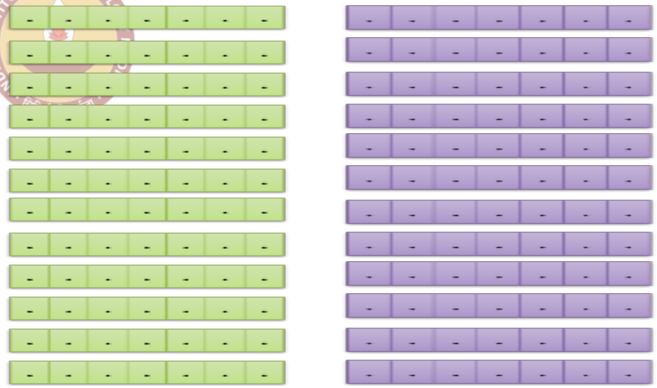
Outline

- Related Work
- Proposed Model
- Analysis of the model
- Application to cross language learning
- **Beyond 2 views – Bridge CorrNet**
- Results

So far we have assumed that there are two views and parallel data exists between them

What if this is not the case?

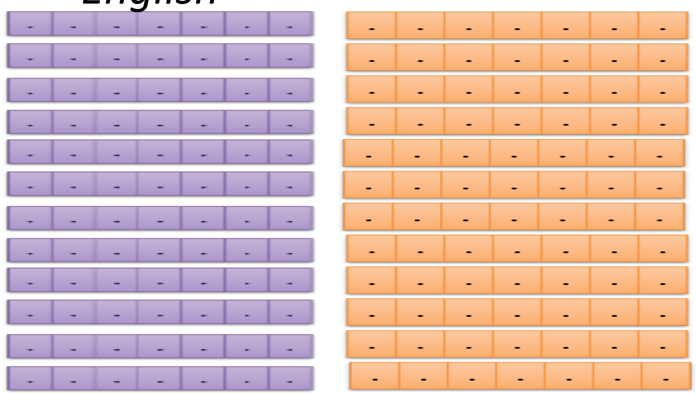
What if there are n views with parallel data missing between many of them?



Parallel data is available between English and French

French English

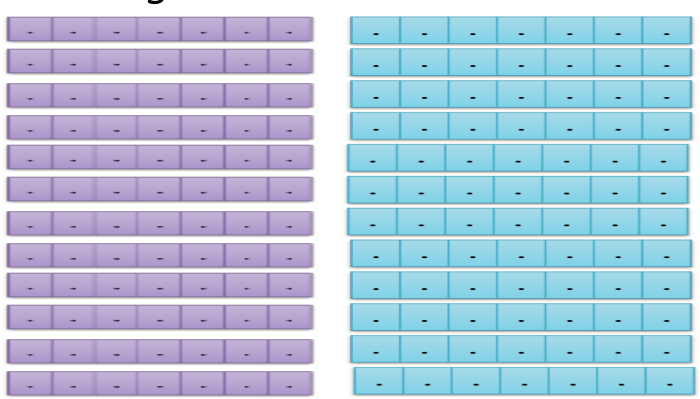
and between English & German



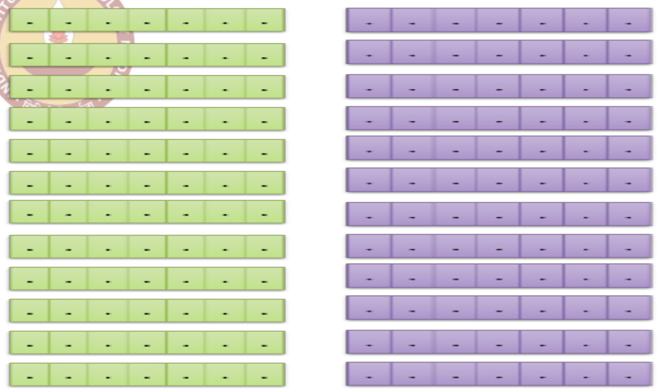
Can we learn a common representation for English, French, German?

English German

and between English & Spanish



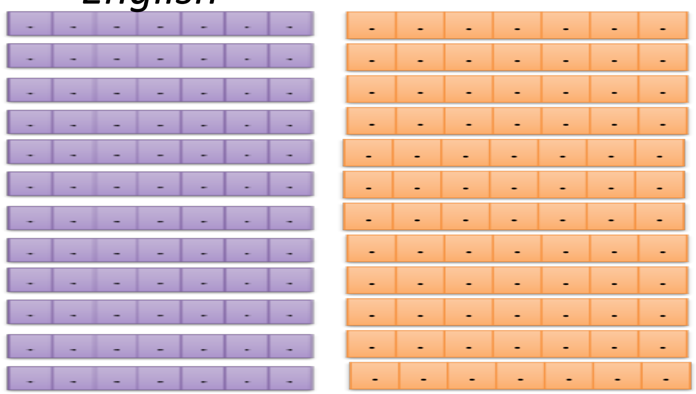
English Spanish



French *English*

Parallel data is available between English and French

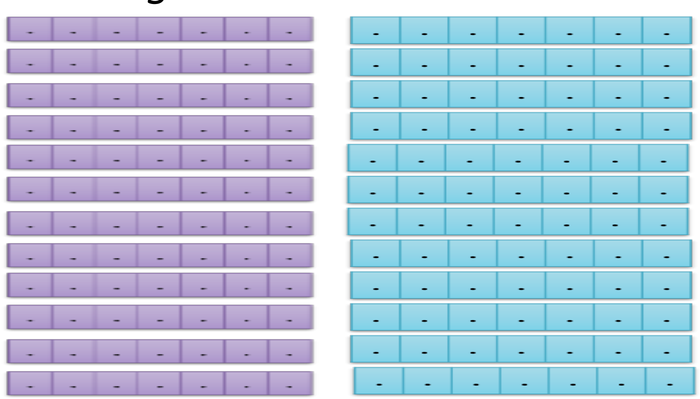
and between English & German



English *German*

Can we learn a common representation for English, French, German & Spanish?

and between English & Spanish



English *Spanish*

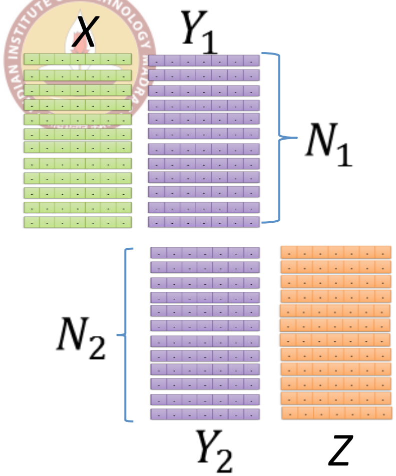
and so on.....

Given n views with parallel data available only between each of these views and a pivot view, can we learn a common representation for all views?

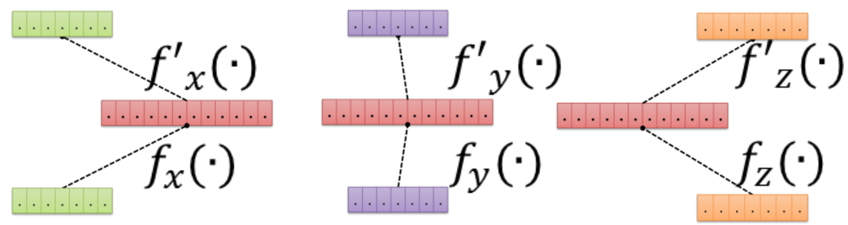
Yes, we can!

(Bridge Correlational Neural Networks)…..

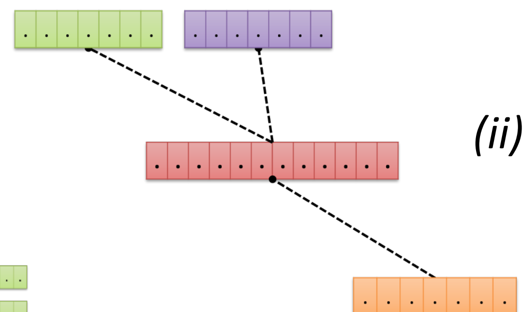
Recall



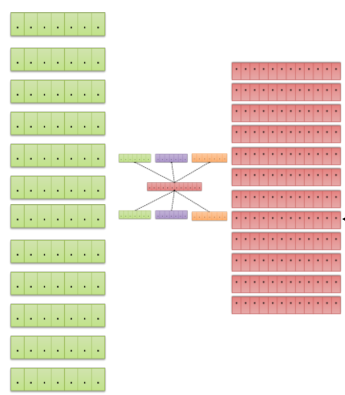
for each view x, y, z find non-linear projections $f_x(\cdot), f_y(\cdot), f_z(\cdot)$ such that



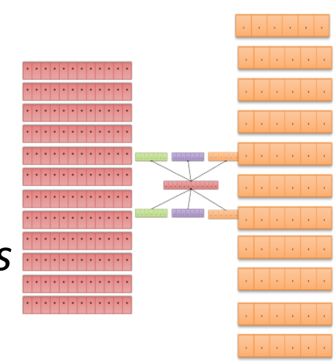
(i) minimize self reconstruction error



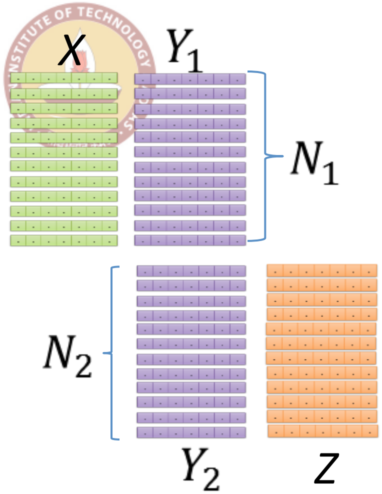
(ii) minimize cross reconstruction error

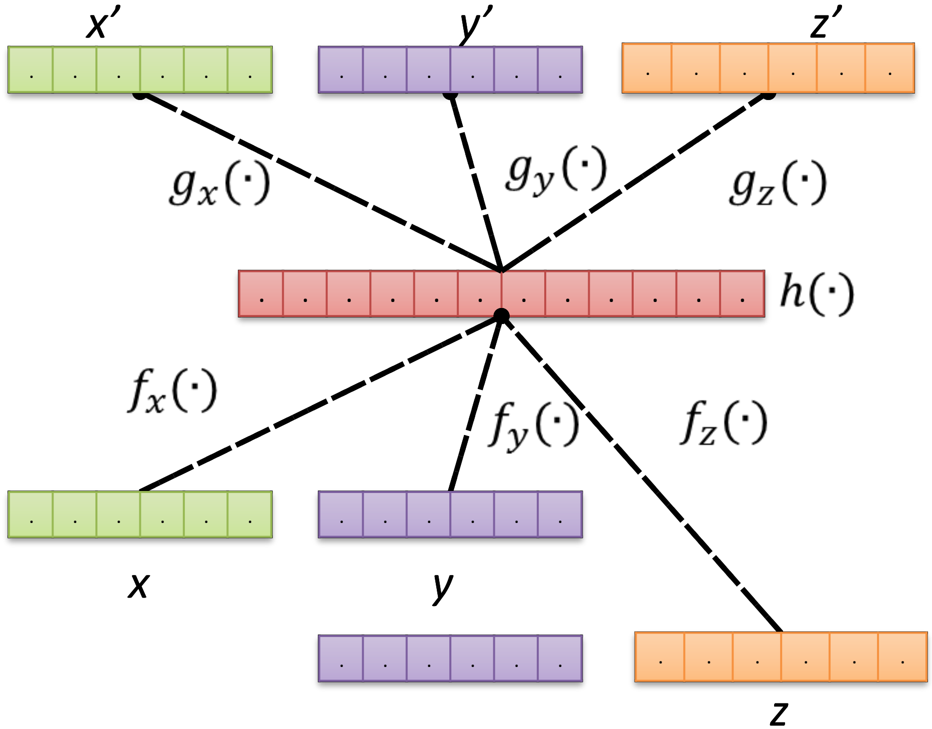


(iii)High correlation even for non-pivot views



Optimize to learn projection matrices W_x, W_y, W_z





$$h(x) = f_x(x) = (\boldsymbol{W_x} x + b)$$
$$h(y) = f_y(y) = (\boldsymbol{W_y} y + b)$$
$$h(z) = f_z(z) = (\boldsymbol{W_z} z + b)$$

$$x' = g_x(h(\cdot)) = (\boldsymbol{W'_x} h(\cdot) + b')$$
$$y' = g_y(h(\cdot)) = (\boldsymbol{W'_y} h(\cdot) + b')$$
$$z' = g_z(h(\cdot)) = (\boldsymbol{W'_z} h(\cdot) + b')$$

$minimize_{(w.r.t.\ W)}$

$$\sum_{i=1}^{N_1} (g_x(f_x(x_i)) - x_i)^2 + \sum_{i=1}^{N_1} (g_y(f_x(x_i)) - y_i)^2 + \sum_{i=1}^{N_1} (g_y(f_y(y_i)) - y_i)^2 + \sum_{i=1}^{N_1} (g_x(f_y(y_i)) - x_i)^2$$

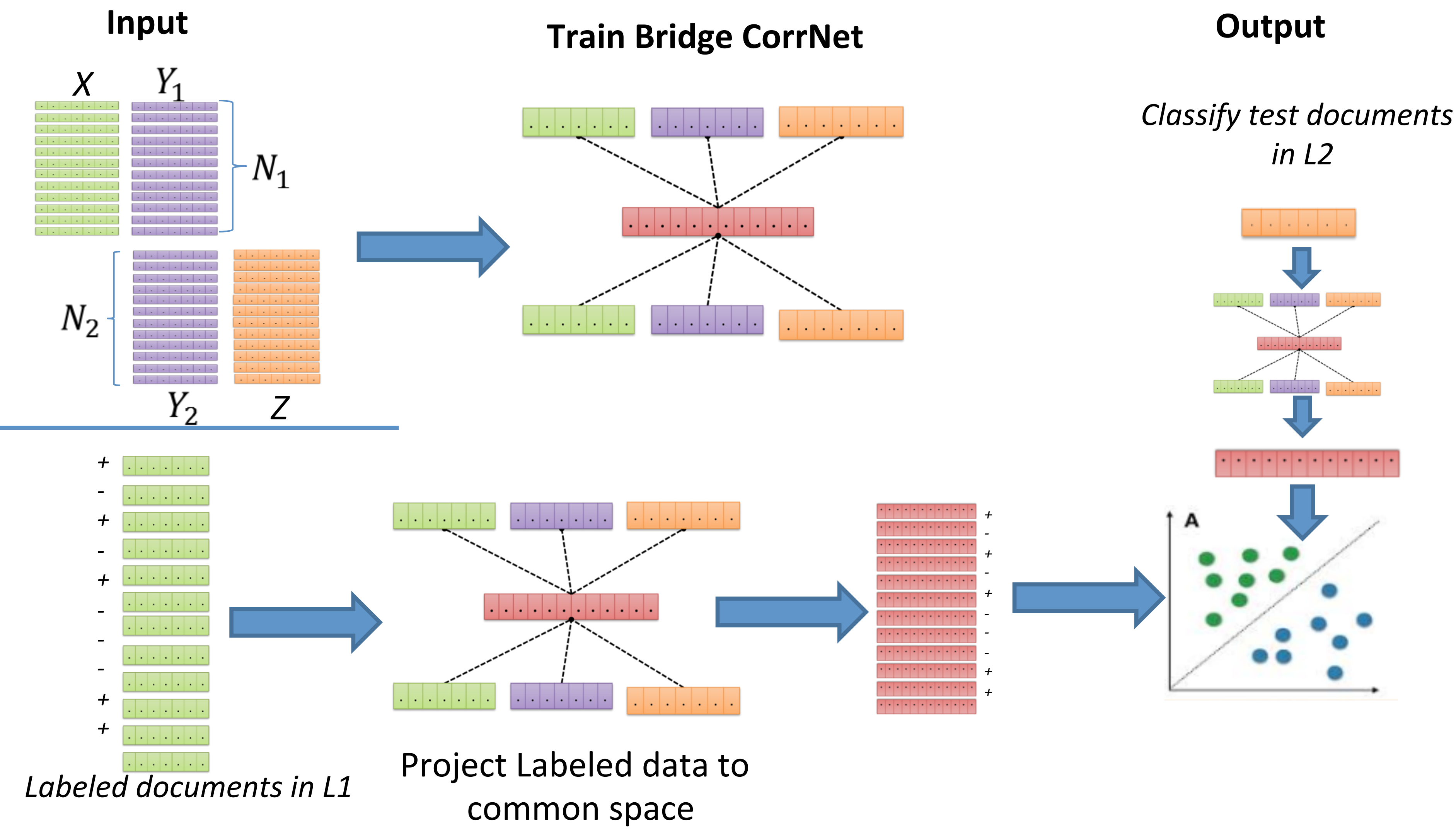
$$+ \sum_{i=1}^{N_2} (g_z(f_z(z_i)) - z_i)^2 + \sum_{i=1}^{N_2} (g_y(f_z(z_i)) - y_i)^2 + \sum_{i=1}^{N_2} (g_y(f_y(y_i)) - y_i)^2 + \sum_{i=1}^{N_2} (g_z(f_y(y_i)) - z_i)^2$$

$$-corr(h(\bar{X}), h(\bar{Y})) - corr(h(\bar{Z}), h(\bar{Y}))$$

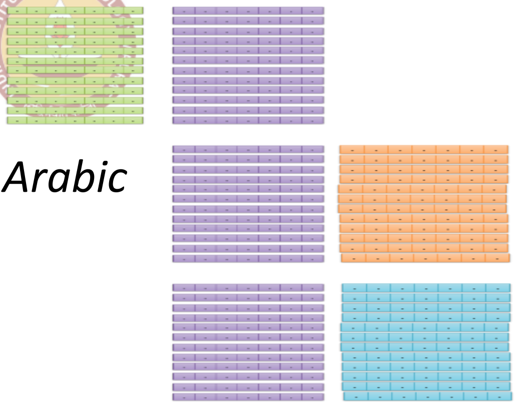
Outline

- Related Work
- Proposed Model
- Analysis of the model
- Application to cross language learning
- Beyond 2 views – Bridge CorrNet
- **Results**

Cross Language Learning



Multilingual TED corpus



- One pivot language - English
- 11 non-pivot languages
- All 11 languages have labeled train/valid/test documents
- Cross language learning experiments with $_{11}C_2$ source-target language pairs

Comparison with Hermann and Blunsom(2014)

	Arabic	German	Spanish	French	Italian	Dutch	Polish	Portu-guese	Rom-anian	Rus-sian	Turkish
Arabic											
German											
Spanish											
French											
Italian											
Dutch											
Polish											
Portuguese											
Romanian											
Russian											
Turkish											

Comparison with Hermann and Blunsom(2014)

	Arabic	German	Spanish	French	Italian	Dutch	Polish	Portu-guese	Rom-anian	Rus-sian	Turkish
Arabic											
German	55.25										
Spanish											
French											
Italian											
Dutch											
Polish											
Portuguese											
Romanian											
Russian											
Turkish								-6.96			

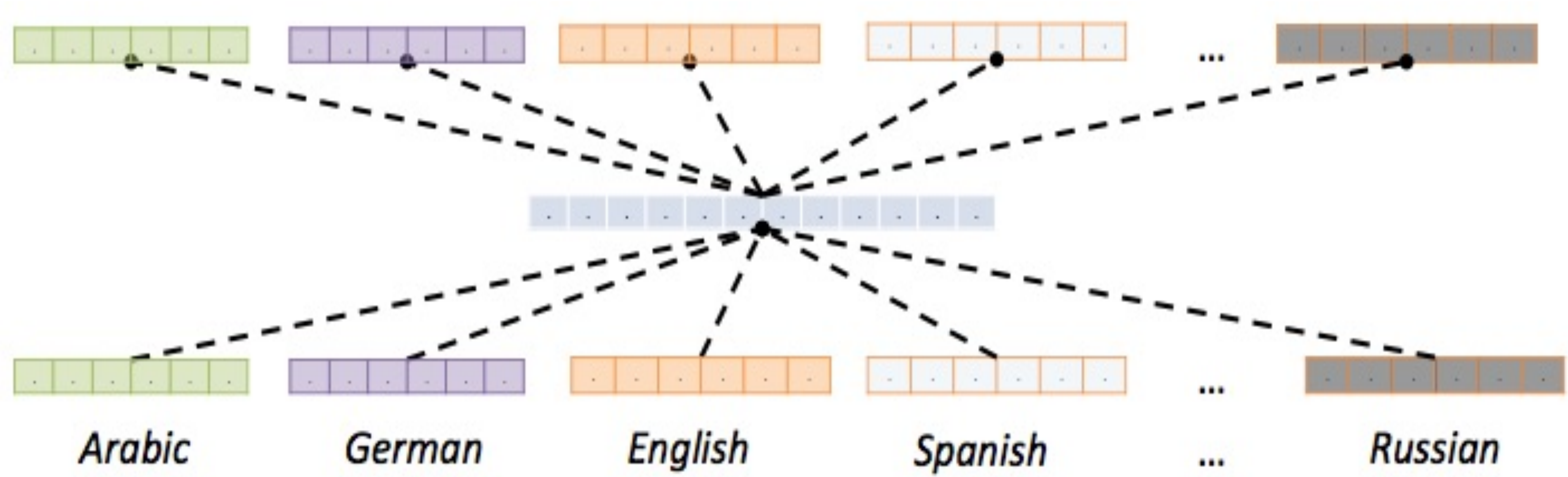
Comparison with Hermann and Blunsom(2014)

	Arabic	German	Spanish	French	Italian	Dutch	Polish	Portu-guese	Rom-anian	Rus-sian	Turkish
Arabic	-20.28	28.35	21.85	21.27	21.93	21.62	23.73	20.27	20.99	16.08	24.69
German	55.25	-1.47	6.96	4.54	19.05	23.18	25.57	8.98	13.64	7.87	15.38
Spanish	31.31	10.97	0.86	12.66	7.27	6.65	12.16	12.80	14.87	3.59	5.15
French	37.78	21.93	19.18	-0.29	20.74	21.70	26.84	23.16	21.70	19.81	18.01
Italian	30.28	1.58	-0.47	-0.87	2.75	13.67	10.26	14.35	1.39	1.99	14.69
Dutch	46.95	1.65	4.75	0.76	14.01	4.35	13.06	10.29	4.26	6.32	12.10
Polish	17.40	0.11	10.67	9.12	10.62	1.30	6.28	4.50	8.70	7.49	0.49
Portuguese	21.82	10.98	9.55	10.56	6.19	7.66	14.19	7.01	10.02	2.05	3.16
Romanian	15.66	13.96	8.59	5.41	6.11	5.01	5.94	5.23	-1.86	4.18	5.58
Russian	38.26	10.76	4.50	1.03	12.88	4.66	12.00	5.93	3.44	-1.84	3.72
Turkish	57.41	2.14	0.05	1.92	11.54	12.13	14.56	-6.96	7.83	0.14	-3.57

Nearest cross language neighbors

English word	Spanish	French	Italian	Dutch	Polish	Pt-Br	Rom'n	Russian	Turkish
					Languages				
market	mercado	marché	mercato	markt	mercado	rynku	piaţă	рынок	pazar
	market	market	market	arbeidsmarkt	lançadas	rynek	piaţă	рынка	piyasa
	place	boursier	vendita	marktonderzoek	mercados	giełdzie	piaţa	рынке	pazara
	comercializan	marketer	azionario	marktaandeel	timbuktu	targ	piata	рыночные	pazarın
oil	petróleo	pétrole	petrolio	olie	petróleo	ropy	petrol	нефти	petrol
	aceite	l'huile	olio	olieprijs	óleo	ropa	petrolul	нефть	petrolün
	petroleros	d'huile	l'olio	olieprijzen	azeite	ropę	petrolului	нефтью	petrolü
	crudo	pétrolières	dell'olio	olieramp	derramamento	oleju	ulei	масло	petrolden
home	casa	maison	casa	thuis	casa	domu	acasă	домой	eve
	hogar	foyer	tramandare	huis	lar	dom	acasa	дома	evde
	casas	domicile	dimora	woning	casas	domem	casă	дом	ev
	hogares	rentre	casetta	thuisblijven	lares	domach	casa	доме	evine
history	historia	l'histoire	storia	geschiedenis	história	historii	istoria	истории	tarih
	historial	d'histoire	qualcuno	wereldgeschiedenis	histórico	historię	istorie	историю	tarihin
	contado	histoire	dell'umanità	history	historia	historia	istoriei	история	tarihi
	participó	historique	popolo	historie	britannica	dziejach	poveste	историей	tarihinde

Correction: We actually build one single model…

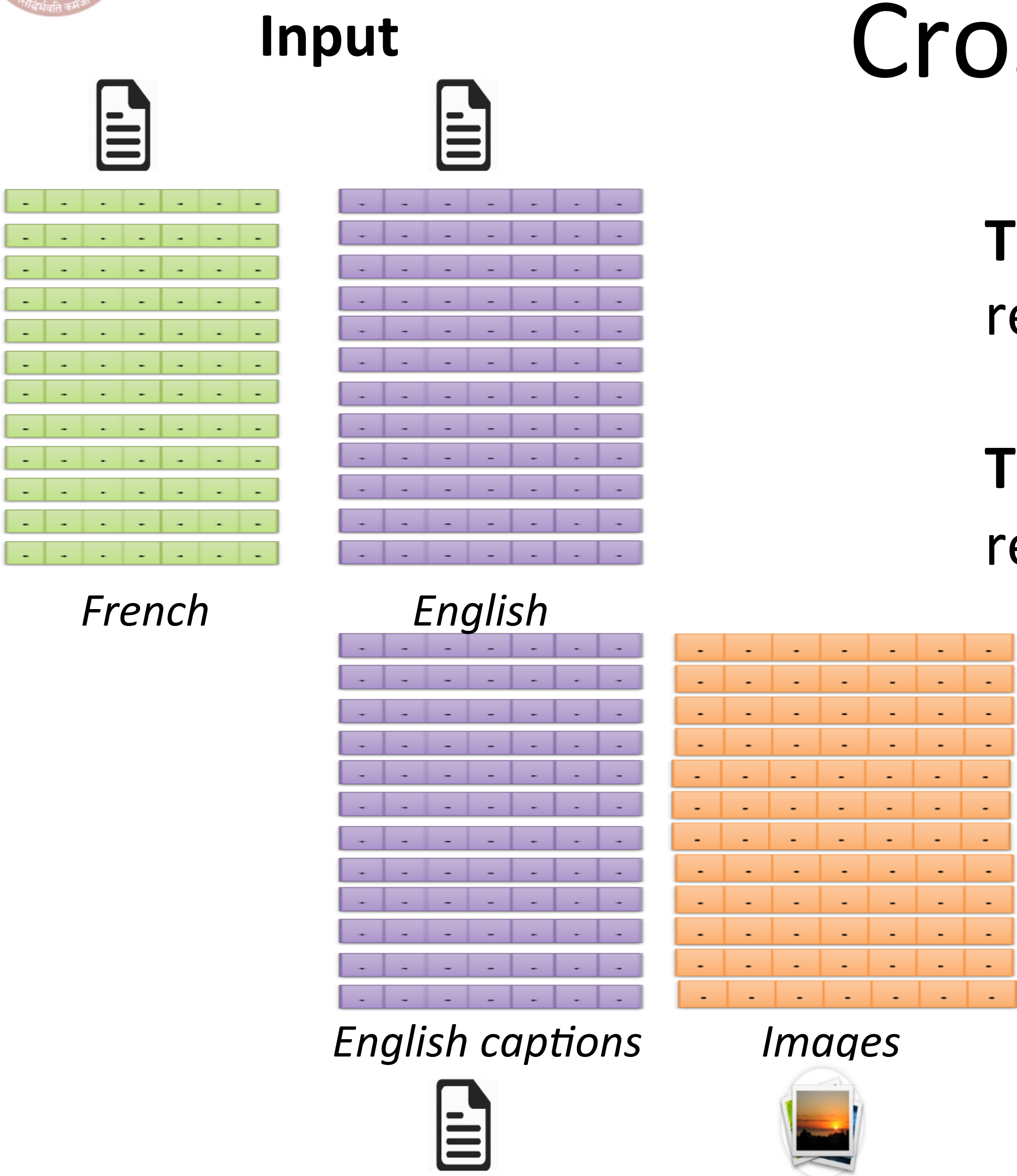


…instead of $_{n}C_{2}$ pairwise Bridge CorrNets…

What about doing this across modalities?

Say, languages and images?

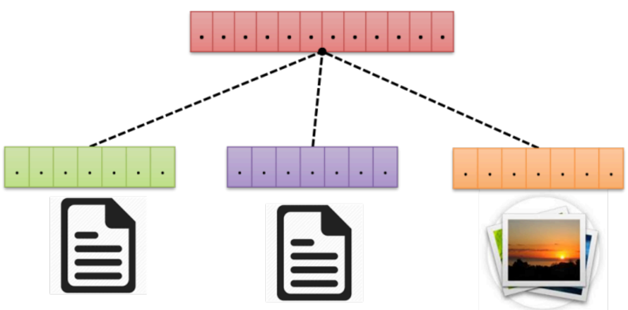
Cross modal access



Task 1: Given a French caption retrieve relevant images

Task 2: Given an image retrieve relevant French captions

Solution



Learn a common representation for English, French and Images using Bridge CorrNet

1. Project given image (query) into this common space

2. Project all French captions (documents) into this common space

3. Rank nearest neighbors based on Euclidean distance

Data

Task 1: 1000 images (queries) and 5000 French/German captions (documents)

Task 2: 5000 French/German captions (queries) and 1000 images (documents)

Results

Model	Captions	Recall @50	
		ImageToCaption	CaptionToImage

Results

Model	Captions	Recall @50	
		ImageToCaption	CaptionToImage
En-Image CorrNet	English	0.452	0.532

Results

Model	Captions	Recall @50	
		ImageToCaption	CaptionToImage
En-Image CorrNet	English	0.452	0.532
BridgeMAE	French	0.069	0.063

Results

Model	Captions	Recall @50	
		ImageToCaption	CaptionToImage
En-Image CorrNet	English	0.452	0.532
BridgeMAE	French	0.069	0.063
2-CorrNet	French	0.109	0.205

Results

Model	Captions	Recall @50	
		ImageToCaption	CaptionToImage
En-Image CorrNet	English	0.452	0.532
BridgeMAE	French	0.069	0.063
2-CorrNet	French	0.109	0.205
Bridge CorrNet	French	0.327	0.232

Results

Model	Captions	Recall @50	
		ImageToCaption	CaptionToImage
En-Image CorrNet	English	0.452	0.532
BridgeMAE	French	0.069	0.063
2-CorrNet	French	0.109	0.205
Bridge CorrNet	French	0.327	0.232
CorrNet + MT	French	0.430	0.416

Results

Model	Captions	Recall @50	
		ImageToCaption	CaptionToImage
En-Image CorrNet	English	0.452	0.532
BridgeMAE	French	0.069	0.063
2-CorrNet	French	0.109	0.205
Bridge CorrNet	French	0.327	0.232
CorrNet + MT	French	0.430	0.416
BridgeMAE	German	0.053	0.058
2-CorrNet	German	0.065	0.098
Bridge CorrNet	German	0.298	0.183
CorrNet + MT	German	0.431	0.343
Random		0.050	0.047

Query

Top-5 nearest German captions

Zwei Pferde stehen auf einem sandigen Strand nahe dem Ocean.
(Two **horses** standing on a sandy beach near the **ocean**)

grasende Pferde auf einer trockenen Weide bei einem Flughafen. (Horses grazing in a dry pasture by an airport.)

ein Elefant , Wasser aufseinen Rückend sprühend , in einem staubigen Bereich neben einem Baum.
(A **elephant** spraying water on its back in a dirt area next to tree .)

ein braunes pferd ißt hohes gras neben einem behälter mit wasser. (Brown horses eating tall grass beside a body of water .)

vier Pferde grasen auf ein Feld mit braunem gras.
(Four **horses** are grazing through a field of **brown grass**.)

Query

Top-5 nearest French captions

Un homme portant une batte de baseball à deux mains lors d'un jeu de balle professionnel.
(**A man holding a baseball bat with two hands** at a professional ball game.)

un joueur de tennis balance une raquette à une balle.
(A tennis player **swinging a racket** at a ball.)

un garçon qui est de frapper une balle avec une batte de baseball.
(A boy that is hitting a ball with a **baseball bat.**)

une équipe de joueurs de baseball jouant un jeu de base-ball.
(A team of **baseball players playing** a game of **baseball**.)

un garçon se prépare à frapper une balle de tennis avec une raquette.
(A boy prepares to hit a tennis ball with a **racquet**.)

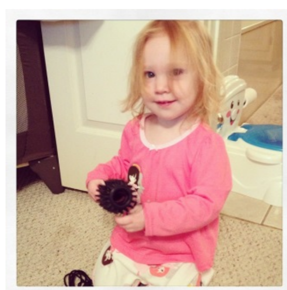
German Query

Speisen und Getränke auf einem Tisch mit einer Frau essen im Hintergrund.

(Food and beverages set on a table with a woman eating in the background .)

Top-5 nearest images

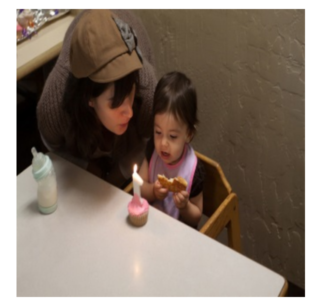
a woman is seen eating at an outdoor table with a bowl of soup and a roll and a beer in the foreground in front of her



a girl playing with a toy on a room floor .

a woman sitting in front of a cake holding her hands

a man and woman cutting a cake with a large knife .



a baby girl sitting in front of a small cupcake with a candle .

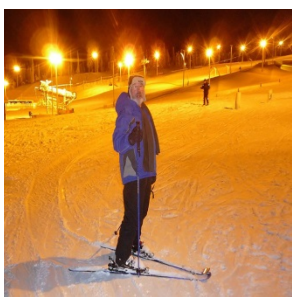
French Query

de ski en se tenant debout dans la neige.

(People wearing ski equipment while standing in snow.)

Top-5 nearest images

two people wearing skiis standing next to each other



a man that is on skis standing in the snow .

a toddler is on a small surfboard at the shore .

a man riding a snow board down a snow covered slope .

a man riding a snow board down a snow covered slope

Thank You!!

- Sarath Chandar, A. P., Khapra, M., Larochelle, H., and Ravindran, B. (2016) "Correlational Neural Networks". In Neural Computation, Vol 28, No. 2, pp. 257-285. MIT Press.
- Rajendran, J., Khapra, M. M., Chandar, S., Ravindran, B. (2016) "Bridge Correlational Neural Networks for Multilingual Multimodal Representation Learning". To appear in the Proceedings of the Fifteenth NAACL:HLT.
- Code: https://github.com/apsarath/CorrNet

http://www.cse.iitm.ac.in/~ravi

http://cmsrv.iitm.ac.in/ilds/home