The William Davidson Faculty of
Industrial Engineering and Management

TECHNION
Israel Institute
of Technology

Bridging Languages Through Images with Deep Partial Canonical Correlation Analysis

GUY ROTMAN [1], IVAN VULIĆ [2] & ROI REICHART [1]

[1] Faculty of Industrial Engineering and Management, Technion, IIT
[2] Language Technology Lab, University of Cambridge

ACL 2018

Motivation

Motivation

- *A visual scene can be described in any language*
  - *Imagine that you are sitting in a restaurant in a foreign country and you need a spoon …*

Goal

- *Find a shared space for textual inputs from several languages*
- *Utilize mutual images to bridge between the textual inputs*

__English__

A man is sitting
at a table
holding a spoon

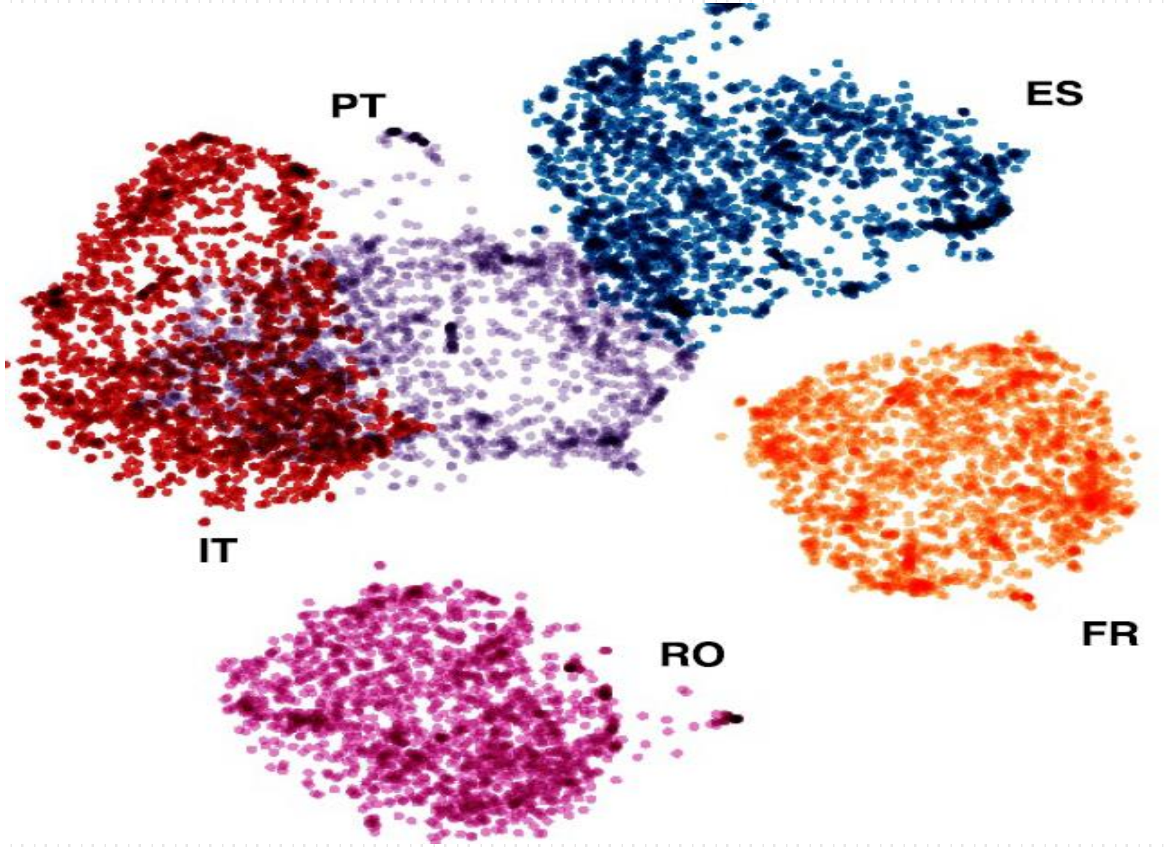
__Spanish__

Un hombre está sentado
en una mesa
sujetando una cuchara

Technical Details

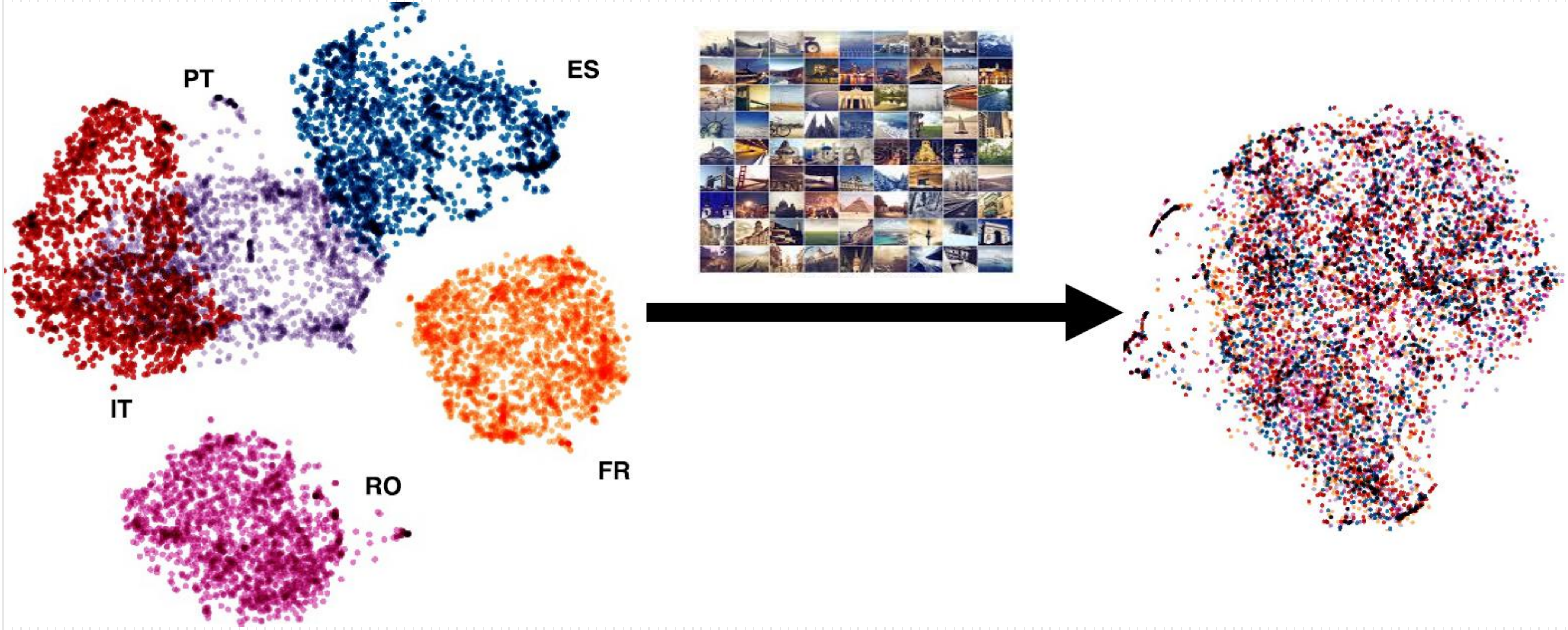
Multilingual Word Embeddings

- *Vectors in different languages are in different spaces*



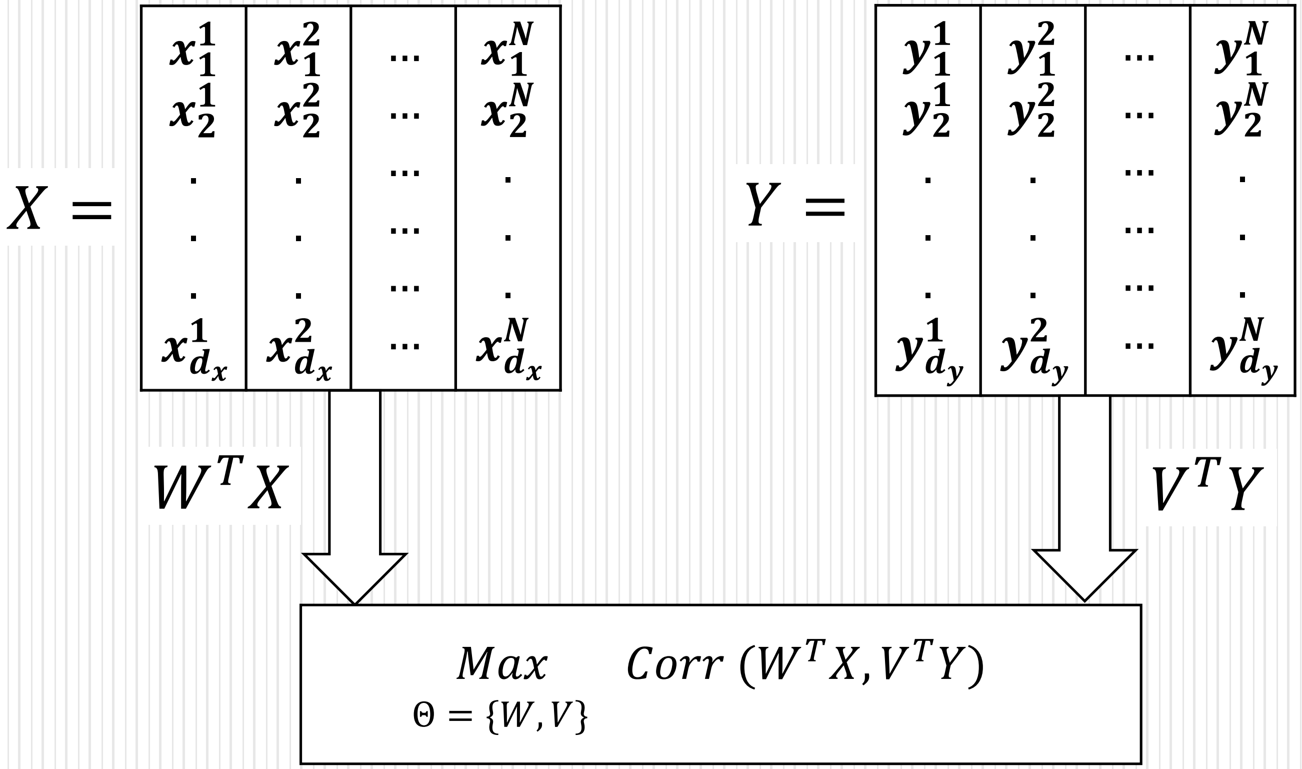
Multilingual Word Embeddings

- *Vectors in different languages are in different spaces*



Mapping Two Views To a Shared Space: Canonical Correlation Analysis (CCA)

- *CCA (Hotelling, 1936) is a statistical technique for finding linear projections of two random matrices such that their projected columns are maximally correlated*



$$X = \begin{array}{|c|c|c|c|} \hline \boldsymbol{x_1^1} & \boldsymbol{x_1^2} & \cdots & \boldsymbol{x_1^N} \\ \boldsymbol{x_2^1} & \boldsymbol{x_2^2} & \cdots & \boldsymbol{x_2^N} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \boldsymbol{x_{d_x}^1} & \boldsymbol{x_{d_x}^2} & \cdots & \boldsymbol{x_{d_x}^N} \\ \hline \end{array} \qquad Y = \begin{array}{|c|c|c|c|} \hline \boldsymbol{y_1^1} & \boldsymbol{y_1^2} & \cdots & \boldsymbol{y_1^N} \\ \boldsymbol{y_2^1} & \boldsymbol{y_2^2} & \cdots & \boldsymbol{y_2^N} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \boldsymbol{y_{d_y}^1} & \boldsymbol{y_{d_y}^2} & \cdots & \boldsymbol{y_{d_y}^N} \\ \hline \end{array}$$

$W^T X$

$V^T Y$

$$\underset{\Theta = \{W, V\}}{Max} \quad Corr\,(W^T X, V^T Y)$$

Mapping Two Views To a Shared Space: Canonical Correlation Analysis (CCA)

- $Objective\ in\ matrix\ form$:

$$\min_{\theta = \{W,V\}} \quad \frac{1}{N-1} ||W^T X - V^T Y||_F^2$$

$$Subject\ to \qquad W^T \hat{\Sigma}_{XX} W \ = V^T \hat{\Sigma}_{YY} V = I$$

- $\hat{\Sigma}_{XY} = \frac{1}{N-1} XY^T,\ \hat{\Sigma}_{XX} = \frac{1}{N-1} XX^T,\ \hat{\Sigma}_{YY} = \frac{1}{N-1} YY^T$
- $X, Y\ have\ zero - mean$

Limitations of CCA

- *Projection is linear*
- *Inapplicable for large datasets due to whitening constraints*:
  - *Hard to compute stochastic estimations of the covariance matrices*
  - *Objective does not decompose over samples*
- *Cannot benefit from an additional view* (*such as images*)

Partial CCA (PCCA)

- *PCCA (Rao, 1969) is a statistical technique for finding linear maximal correlated projections of two random matrices* ***conditioned on a third variable***

$$\underset{\Theta=\{W,V\}}{Max} \quad Corr\,(W^T(X|Z), V^T(Y|Z))$$

- *Z (a visaal input) is a mutual variable of X and Y (textual inputs)*
- *PCCA was not used before in the multilingual multimodal setup*

New model - Deep Partial CCA (DPCCA)

- *CCA has a deep variant − Deep CCA (Andrew et al.,* 2013)

New model - Deep Partial CCA (DPCCA)

- *CCA has a deep variant − Deep CCA (Andrew et al.,* 2013)

- *Can we develop a deep variant for Partial CCA?*

    - *Partial CCA suffers from similar limitations to those of CCA*

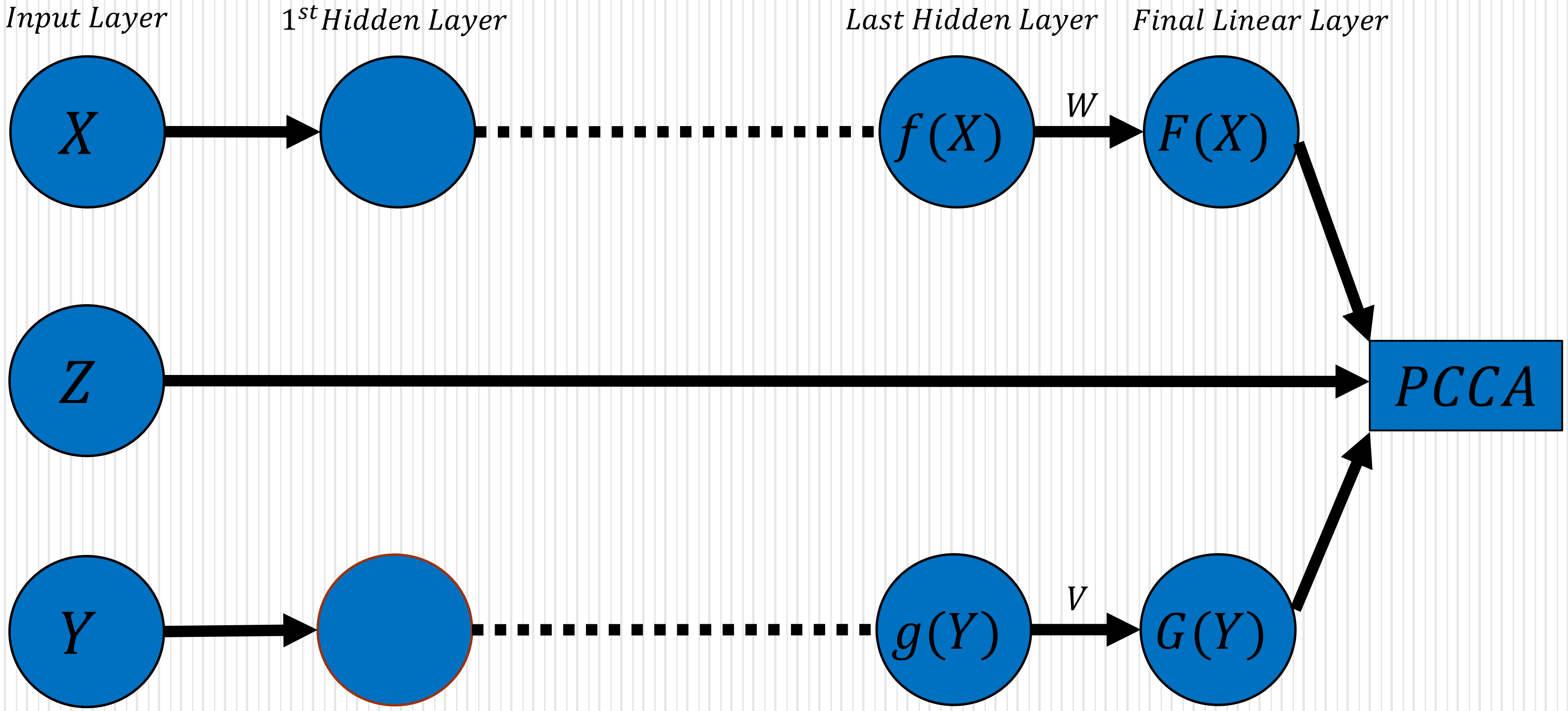
    - *A new stochastic optimization algorithm is required*

The DPCCA Model

Architecture of Deep Partial CCA (DPCCA) - Variant A

A man is sitting at a table holding a spoon

Un hombre está sentado en una mesa sujetando una cuchara



Input Layer

1^{st} *Hidden Layer*

Last Hidden Layer

Final Linear Layer

X

$f(X)$

W

$F(X)$

Z

$PCCA$

Y

$g(Y)$

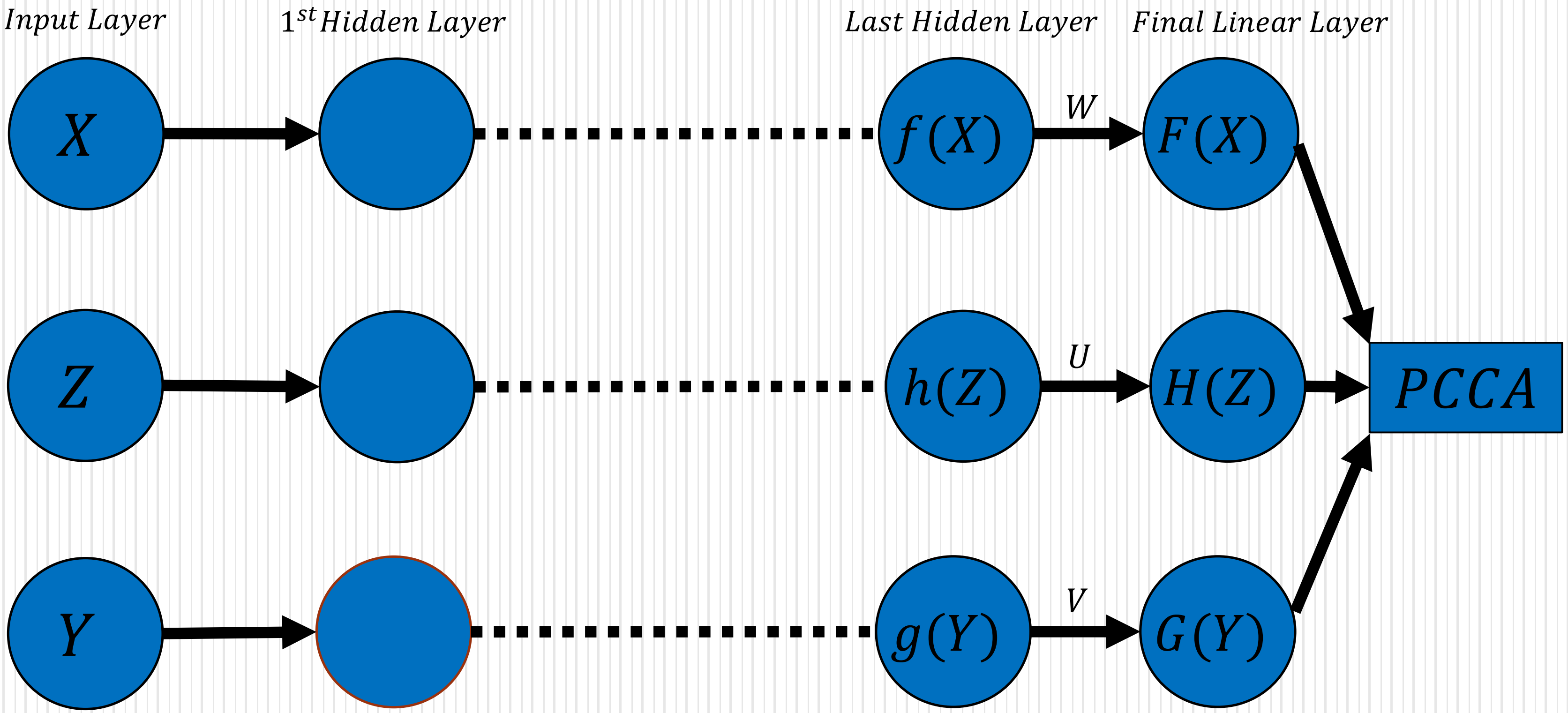
V

$G(Y)$

Architecture of Deep Partial CCA (DPCCA) - Variant B

A man is sitting
at a table
holding a spoon

Un hombre está sentado
en una mesa
sujetando una cuchara



Deep Partial CCA (DPCCA)

- (1) *learn non-linear representations of* X *and* Y:

  $$F(X) = W^T f(X), \qquad G(Y) = V^T g(Y)$$

  - f *and* g *are two deep neural networks*
  - W *and* V *are the final projection matrices*

Deep Partial CCA (DPCCA)

- (2) *perform multivariate linear multiple regressions for* $F(X)$ *and* $G(Y)$ *on a shared variable* Z:

$$F(X) = \underbrace{AZ}_{explained} + \underbrace{F(X|Z)}_{residual} \qquad \min_{A} \frac{1}{N-1} ||F(X) - AZ||_F^2$$

$$\Rightarrow$$

$$G(Y) = \underbrace{BZ}_{explained} + \underbrace{G(Y|Z)}_{residual} \qquad \min_{B} \frac{1}{N-1} ||G(Y) - BZ||_F^2$$

Deep Partial CCA (DPCCA)

- (2) *perform multivariate linear multiple regressions for* $F(X)$ *and* $G(Y)$ *on a shared variable* Z:

$$F(X) = AZ + F(X|Z)$$

$$G(Y) = BZ + G(Y|Z)$$

$$\Longrightarrow$$

$$\min_{A} \frac{1}{N-1} ||F(X) - AZ||_F^2$$

$$\min_{B} \frac{1}{N-1} ||G(Y) - BZ||_F^2$$

- (3) *compute the residual matrices and their covariances w.r.t. the optimal solutions*:

$$F(X|Z) = F(X) - \hat{A}Z \qquad G(Y|Z) = G(Y) - \hat{B}Z$$

$$\hat{\Sigma}_{FF|Z} = \frac{1}{N-1} F(X|Z)F(X|Z)^T \qquad \hat{\Sigma}_{GG|Z} = \frac{1}{N-1} G(Y|Z)G(Y|Z)^T$$

Deep Partial CCA (DPCCA)

- $(4)\ perform\ CCA\ on\ the\ residuals$:

$$\min_{\theta=\{W_f,W,V_g,V\}} \frac{1}{N-1}||F(X|Z)-G(Y|Z)||_F^2$$

$$Subject\ to \qquad \hat{\Sigma}_{FF|Z}=\hat{\Sigma}_{GG|Z}=I$$

Deep Partial CCA (DPCCA) – Optimization

- *Optimization is not trivial*

Deep Partial CCA (DPCCA) – Optimization

- *Optimization is not trivial*

- *We introduce new stochastic optimization algorithms for our DPCCA variants*

- *Full Pseudocode is given in the paper*

Deep Partial CCA (DPCCA) – Optimization

- *Optimization is not trivial*

- *We introduce new stochastic optimization algorithms for our DPCCA variants*

- *We adopt some key techniques from the Nonlinear Orthogonal Iteration* (*NOI*) *algorithm which was suggested for Deep CCA* (*Wang et al.*, 2015)

- *Full Pseudocode is given in the paper*

Experiments and Results

Experimental Setup – Tasks and Datasets

- *First Task*: *Cross-lingual image description retrieval*

English

A man is sitting at a table holding a spoon

Spanish

Un hombre está sentado en una mesa sujetando un tenedor

Un hombre está sentado en una mesa sujetando una cuchara

Un hombre está sentado en un balcon sujetando una cuchara

.

.

.

- *Dataset*: *Multi30k* (*Elliott et al.*, 2016)

Experimental Setup – Tasks and Datasets

- *First Task*: *Cross-lingual image description retrieval*

English

A man is sitting at a table holding a spoon

Spanish

Un hombre está sentado en una mesa sujetando un tenedor

Un hombre está sentado en una mesa sujetando una cuchara

Un hombre está sentado en un balcon sujetando una cuchara

.

.

.

- *Dataset*: *Multi30k* (*Elliott et al.*, 2016)

Experimental Setup – Tasks and Datasets

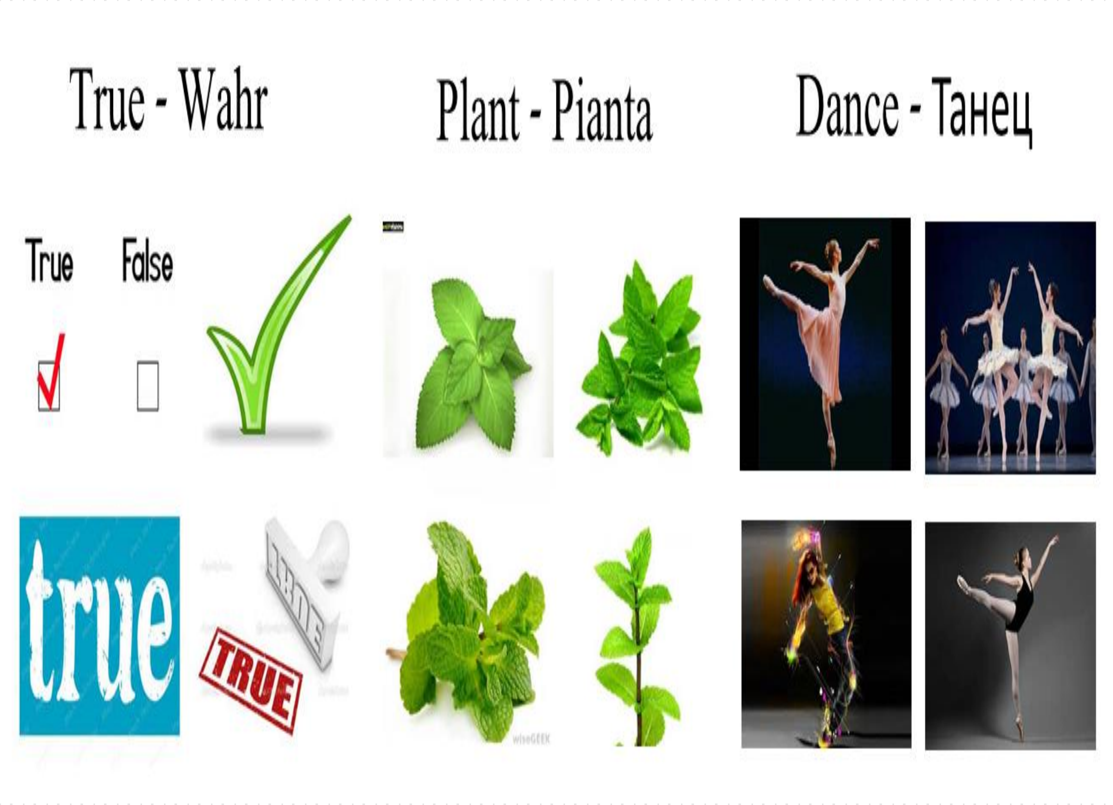
- *Second Task*: *Multilingual Word Similarity*

English		**German**		**Italian**		**Russian**	
inspect-examine	9.2	*prüfen-überprüfen*	9.8	*inspezionare-esaminare*	8.5	осматривать-изучать	5.3
easy-flexible	3.7	*leicht-flexibel*	3.4	*facile-flessibile*	2.5	покладистый-гибкий	4.0
plane-airport	1.6	*flugzeug-flughafen*	5.9	*aereo-aeroporto*	6.2	самолет-аэропорт	1.3

- *Dataset*: *Multilingual Simlex*-999 (*Leviant and Reichart*., 2015)

New Dataset – Word Image Word (WIW)

- *Word pairs in different languages with mutual images*



POS	EN-DE	EN-IT	EN-RU
N	4606	4735	4106
A	405	416	348
V	392	400	227
AVB	167	161	142
PP	12	12	9
TOTAL	5598	5740	4838

- *The new dataset is available at: github.com/rotmanguy/DPCCA*

Experimental Setup - Baselines

- *Linear and deep CCA-based models*:
  - *Probabilistic Partial CCA* (*PPCCA*) (*Mukuta*, 2014) − *T*
  - *Nonparametric CCA (NCCA) (Michaeli et al., 2016) - T*
  - *Generalized CCA* (*GCCA*) (*Horst*, 1961) − *TI*
  - *Deep CCA* (*DCCA*) *with various optimization algorithms* − *T*
  - *Deep CCA Autoencoder* (*DCCAE*) (*Wang et al.*, 2015) − *T*

Text − *T*,   *Text* + *Images* − *TI*

Experimental Setup - Baselines

- *Linear and deep CCA-based models*:
  - *Probabilistic Partial CCA* (*PPCCA*) (*Mukuta*, 2014) − *T*
  - *Nonparametric CCA (NCCA) (Michaeli et al*., 2016*)* - *T*
  - *Generalized CCA* (*GCCA*) (*Horst*, 1961) − *TI*
  - *Deep CCA* (*DCCA*) *with various optimization algorithms* − *T*
  - *Deep CCA Autoencoder* (*DCCAE*) (*Wang et al*., 2015) − *T*

- *Other related works*:
  - *Bridge Correlational Networks* (*BCN*) (*Rajendran et al*., 2016) − *TI*
  - *Image Pivoting* (*Gella et al*., 2017) − *TI*

Text − *T*,   *Text* + *Images* − *TI*

Main Results

- *PCCA gets very good results, outperforming NN based methods and linear methods (including CCA, Image Pivoting, BCN …)*

- *DPCCA is the best model, outperforming all baseline*

- *Training with images improves performance on words that are more abstract, such as adjectives and verbs*

Cross-lingual Image Description Retrieval

Model	English to German	German to English
DPCCA Variant A	83.6%	82.7%
DPCCA Variant B	84.8%	**83.9%**
DPCCA Variant B + DCCA NOI (Concatenation)	**86.3%**	83.7%
DCCA NOI	84.9%	83.0%
IMG PIVOTING	78.9%	78.1%
BCN	62.8%	62.9%
PCCA	**82.4%**	**78.7%**
CCA	80.3%	75.4%
GCCA	74.2%	74.3%

- *Results are reported on BLEU* + 1

Multilingual Word Similarity

Model	EN - ADJ	EN - Verbs	EN - Nouns	DE - ADJ	DE –Verbs	DE - Nouns
DPCCA Variant A	**64.0%**	31.1%	36.9%	43.0%	**32.1%**	**40.4%**
DPCCA Variant B	62.6%	**31.6%**	**38.2%**	**46.2%**	31.9%	39.9%
DCCA NOI	61.1%	30.8%	36.1%	44.1%	29.7%	39.8%
PCCA	61.4%	29.6%	34.0%	30.5%	14.3%	34.0%
CCA	55.7%	29.7%	32.1%	28.4%	15.7%	34.6%
GCCA	63.6%	28.0%	37.8%	44.6%	27.7%	39.8%

- *Results are reported on Spearman's correlation coefficient*

Summary

- *Goal*: *Learning a shared bilingual space for textual inputs*

Summary

- *Goal*: *Learning a shared bilingual space for textual inputs*

- *Our Contributions*:
  - *Method*: *Adding mutual visual information to the learning process*
  - *Model*: *Applying PCCA to our settings, and introducing its deep variants*
  - *Optimization*: *New optimiztion algorithm for DPCCA*
  - *Results*: *Improvements over previous work*
  - *New Dataset*: *Word Image Word (WIW)*

Future Work

- *Exapnding DPCCA to support more than two languages*

- *Exploiting the internal structure of images and sentences*

Thank you!

- *Code and data are available at*:

github.com/rotmanguy/DPCCA