

# Trimming and Improving Skip-thought Vectors

--Representing sentences as vectors

Shuai Tang (唐帅)

Why?

Sentence → Vector

# Why?

Sentence → Vector

We communicate in sentences,  
and they convey our thoughts.

# Why?

Sentence → Vector

Vector is an efficient type of representation  
for machines to operate on.

# Why?

Sentence → Vector

If we convert a sentence into a vector that **captures the meaning** of the sentence, then Google can do much better searches; they can search based on what's being said in a document. (Hinton, 2015)

Natural Reasoning

# Machine Learning

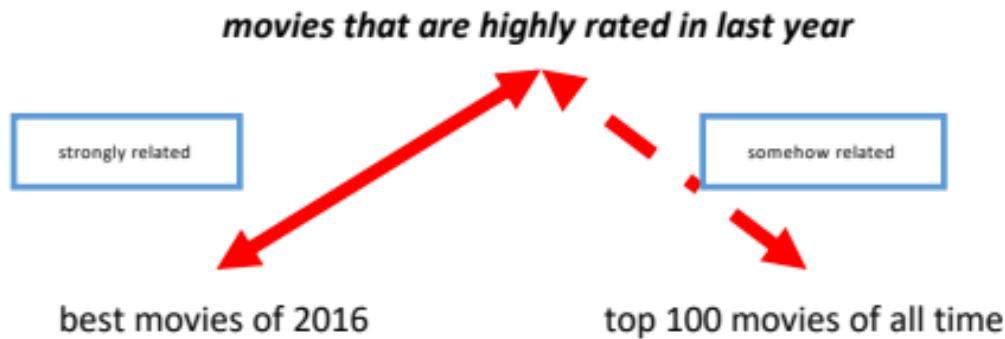
Sentence → Vector

Learn from  
data!

Supervised Learning

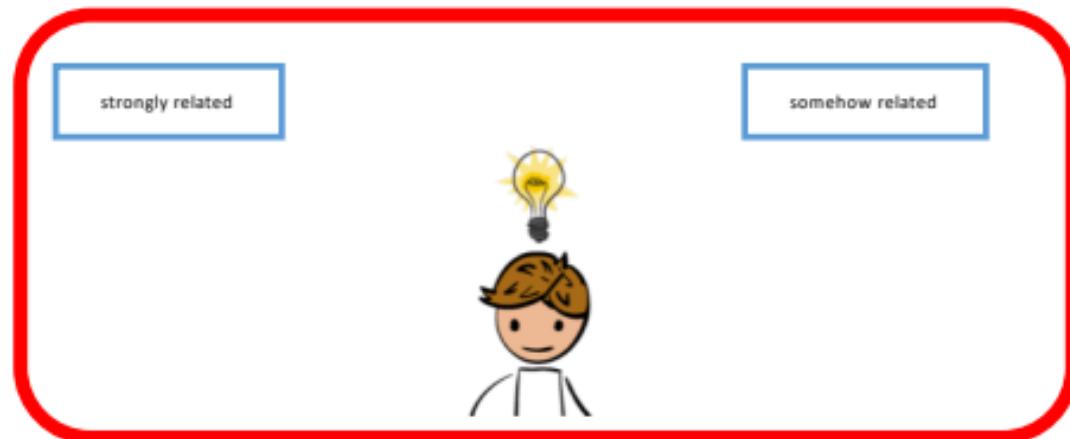
Unsupervised Learning

# Supervised Learning



# Supervised Learning

labels



# Unsupervised Learning

Sentence → Vector

Learn from  
data!  
Without  
labels!

# Existing Models

- Bag-of-words (BOW)
- Continuous Bag-of-words (CBOW)/ Skip-gram
- Sequence to Sequence (Seq2Seq)
- Skip-thought

# Existing Models

- **Bag-of-words (BOW)**
  - Harris, Word1954
- Continuous Bag-of-words (CBOW)/ Skip-gram
- Sequence to Sequence (Seq2Seq)
- Skip-thought

# Bag-of-words model (BOW)

- Corpus
  - i love you.
  - however, you don't love me.
  - it is a sad story.
- Dictionary
  - {i, love, you, however, don't, me, it, is, a, sad, story}
- Representations
  - [1, 1, 1, 0, 0, 0, 0, 0, 0, 0]
  - [0, 0, 1, 1, 1, 0, 0, 0, 0, 0]
  - [0, 0, 0, 0, 0, 0, 1, 1, 1, 1]

# Bag-of-words model (BOW)

- Corpus
  - i love you.
  - however, you don't love me.
  - it is a sad story.
- Dictionary
  - {i, love, you, however, don't, me, it, is, a, sad, story}
- Representations
  - [1, 1, 1, 0, 0, 0, 0, 0, 0, 0]
  - [0, 0, 1, 1, 1, 0, 0, 0, 0, 0]
  - [0, 0, 0, 0, 0, 0, 1, 1, 1, 1]

# Bag-of-words model (BOW)

- Corpus
  - i love you.
  - however, you don't love me.
  - it is a sad story.
- Dictionary
  - {i, love, you, however, don't, me, it, is, a, sad, story}
- Representations
  - [1, 1, 1, 0, 0, 0, 0, 0, 0, 0]
  - [0, 0, 1, 1, 1, 0, 0, 0, 0, 0]
  - [0, 0, 0, 0, 0, 1, 1, 1, 1, 1]

# Bag-of-words model (BOW)

- Corpus
  - i love you.
  - however, you don't love me.
  - it is a sad story.
- Dictionary
  - {i, love, you, however, don't, me, it, is, a, sad, story}
- Representations
  - [1, 1, 1, 0, 0, 0, 0, 0, 0, 0]
  - [0, 0, 1, 1, 1, 0, 0, 0, 0, 0]
  - [0, 0, 0, 0, 0, 1, 1, 1, 1, 1]

Fa  
stl

# Bag-of-words model (BOW)

- Corpus
  - i love you.
  - however, you don't love me.
  - it is a sad story.
- Dictionary
  - {i, love, you, however, don't, me, it, is, a, sad, story}
- Representations
  - [1, 1, 1, 0, 0, 0, 0, 0, 0, 0]
  - [0, 0, 1, 1, 1, 1, 0, 0, 0, 0]
  - [0, 0, 0, 0, 0, 0, 1, 1, 1, 1]

# Existing Models

- Bag-of-words (BOW)
- **Continuous Bag-of-words (CBOW)/ Skip-gram**
  - Mikolov et al., NIPS2013
- Sequence to Sequence (Seq2Seq)
- Skip-thought

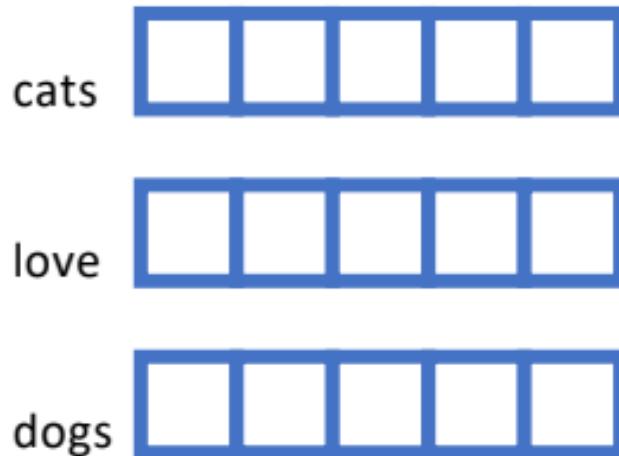
# Continuous BOW/ Skip-gram (Mikolov et al. NIPS2013)

cats

love

dogs

# Continuous BOW/ Skip-gram (Mikolov et al. NIPS2013)



# Continuous BOW/ Skip-gram (Mikolov et al. NIPS2013)

	0	-	-	0	
	.			.	
cats	4	5	0	1	2
	0	-	-	-	-
	.			.	
love	0	1	9	3	4
	0	-	-	-	-
	.			.	
dogs	5	4	1	2	3

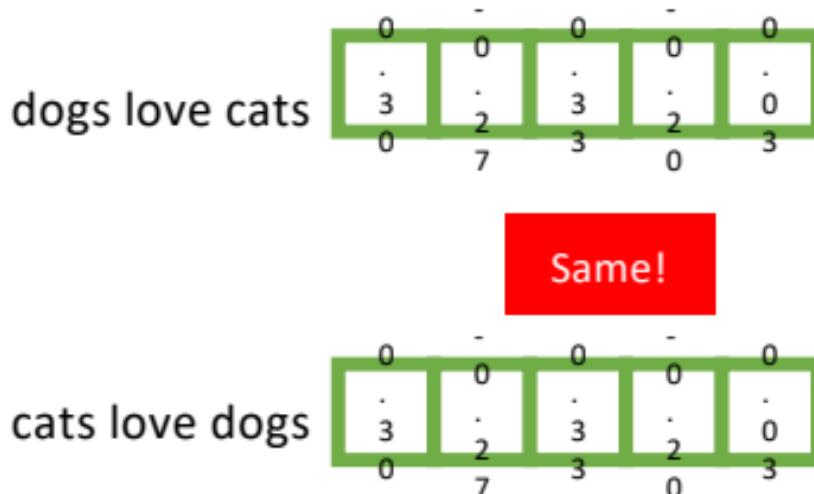
# Continuous BOW/ Skip-gram (Mikolov et al. NIPS2013)

	0	-	-	0	
	.	0	0	.	.
	4	.	5	1	2
cats	0	.	.	.	.
	.	0	0	0	0
	1	.	9	3	4
love	0	-	-	-	0
	.	0	0	0	.
	5	.	4	1	2
dogs	0	-	-	-	0
	.	0	0	0	.
	3	.	2	3	0
cats love dogs	0	-	-	-	0
	0	7	3	0	3

## Continuous BOW/ Skip-gram (Mikolov et al. NIPS2013)

	0	-	-	0
cats	4	5	1	2
	0	-	-	+
love	1	9	3	4
	0	-	-	+
dogs	5	4	1	2
	0	-	-	0
cats love dogs	3	2	3	2
	0	7	3	0
	0	-	-	3

# Continuous BOW/ Skip-gram (Mikolov et al. NIPS2013)

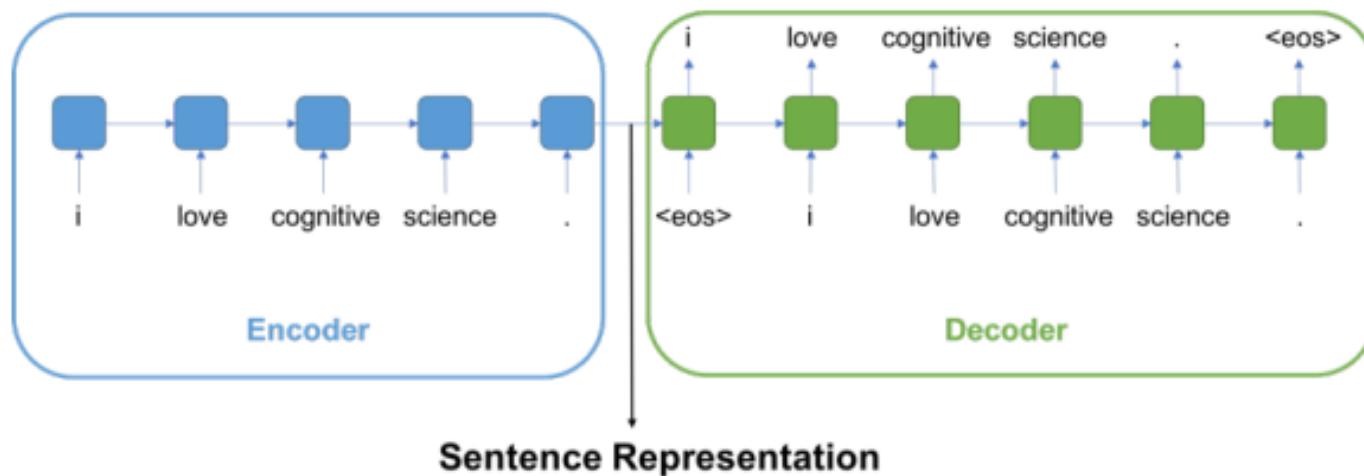


# Existing Models

- Bag-of-words (BOW)
- Continuous Bag-of-words (CBOW)/ Skip-gram
- **Sequence to Sequence (Seq2Seq)**
  - Sutskever, Vinyals & Le (NIPS2014)
  - Dai & Le (NIPS2015)
- Skip-thought

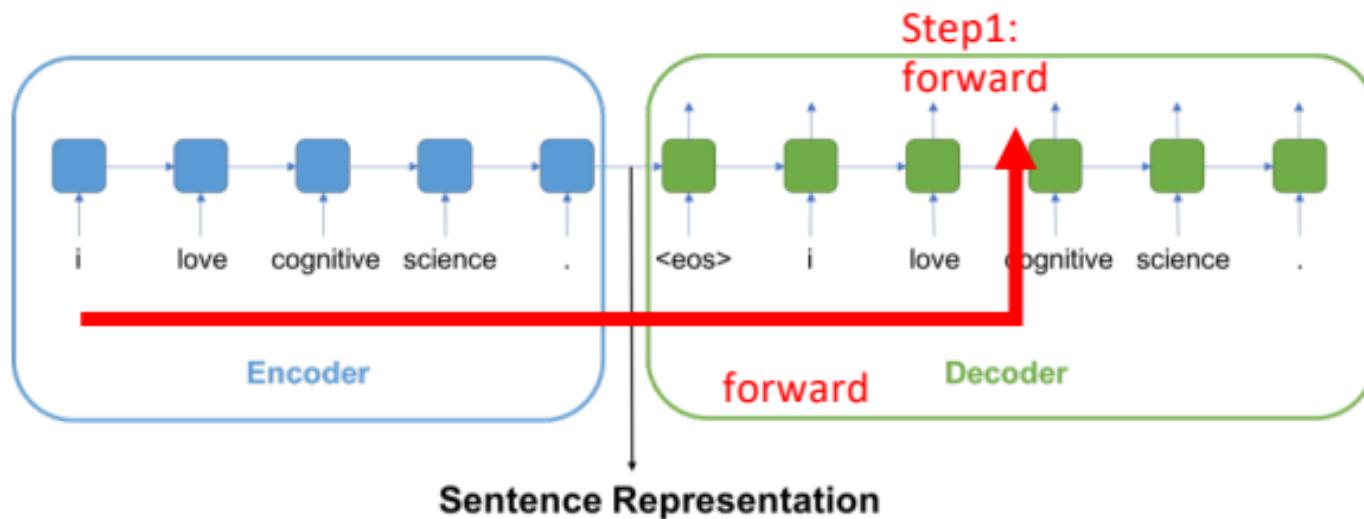
# Sequence to Sequence (Dai & Le, NIPS2015)

- With Recurrent Neural Networks



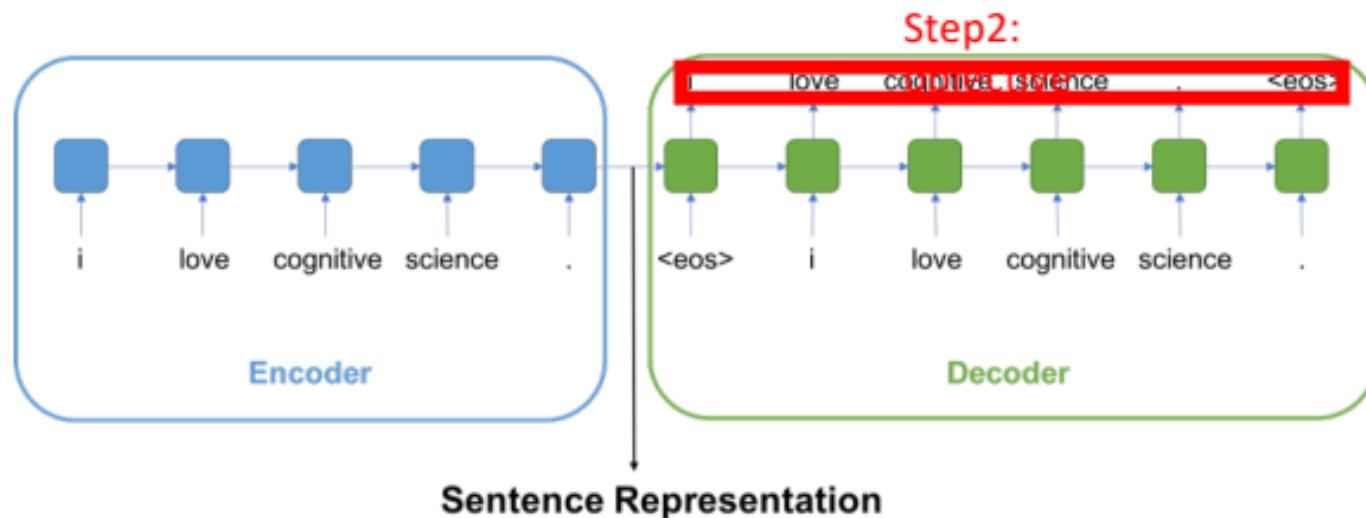
# Sequence to Sequence (Dai & Le, NIPS2015)

- Each training iteration...



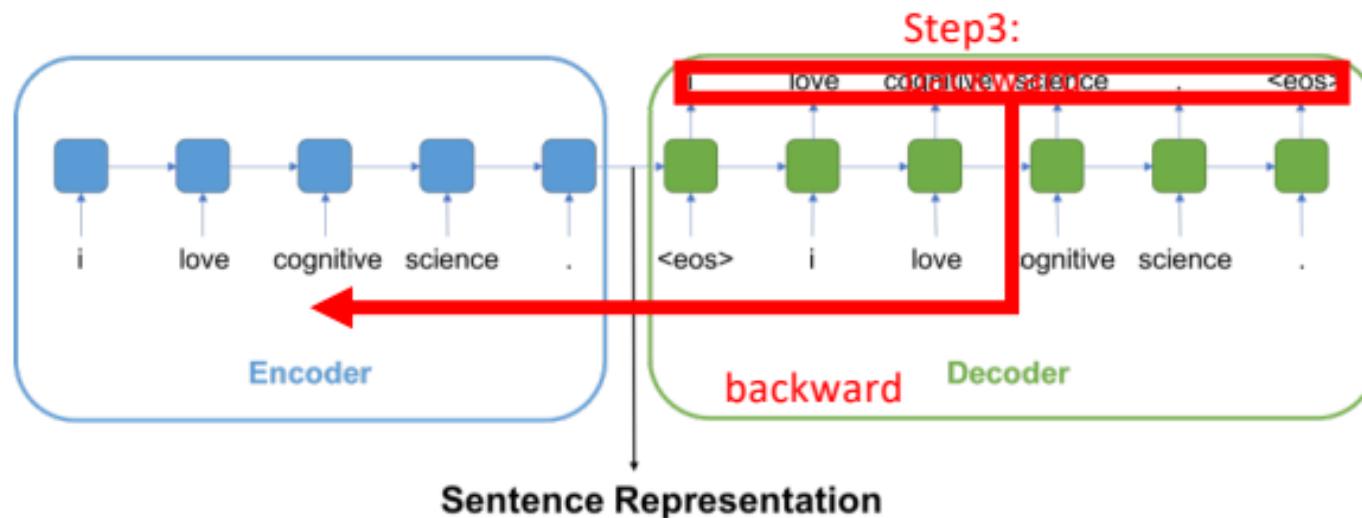
# Sequence to Sequence (Dai & Le, NIPS2015)

- Each training iteration...



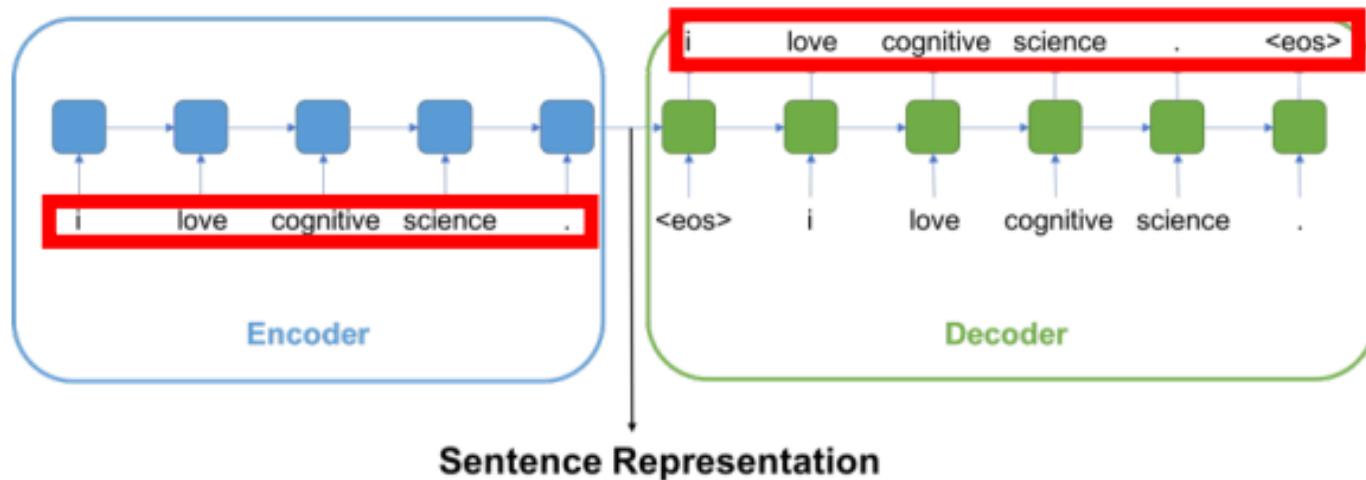
# Sequence to Sequence (Dai & Le, NIPS2015)

- Each training iteration...



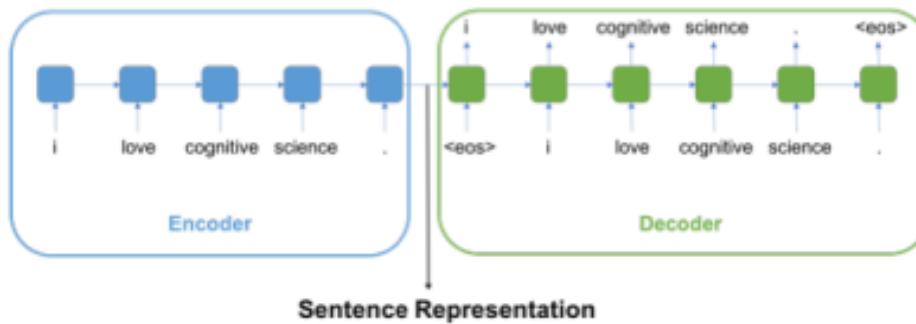
# Sequence to Sequence (Dai & Le, NIPS2015)

- With Recurrent Neural Networks



# Sequence to Sequence (Dai & Le, NIPS2015)

- Pros:
  - Word-order information is utilized in training.
- Cons:
  - Training is slow.



# Existing Models

- Bag-of-words (BOW)
- Continuous Bag-of-words (CBOW)/ Skip-gram
- Sequence to Sequence (Seq2Seq)
- **Skip-thought**
  - Kiros et al., NIPS2015

# Trimming and Improving Skip-thought Vectors

- **Skip-thought**
  - **Kiros et al., NIPS2015**
- Our hypotheses to improve skip-thought
- Comparison between our trimmed skip-thought model and the skip-thought model
- Conclusion

# Skip-thought (Kiros et al., NIPS2015)

- Skip-thought model is for learning a generic sentence **encoder** .

Encoder - Decoder



# Skip-thought (Kiros et al., NIPS2015)

- The skip-thought model learns to encode a sentence, and decode its surrounding two sentences, instead of itself.



The context in which words and sentences are understood plays an important role in human comprehension.  
(Altmann & Steedman, 1988; Binder & Desai, 2011)

# Skip-thought (Kiros et al., NIPS2015)

- The model contains
  - an **encoder**
  - a **previous decoder**
  - a **next decoder**

} 3 parametric functions needs to be learned



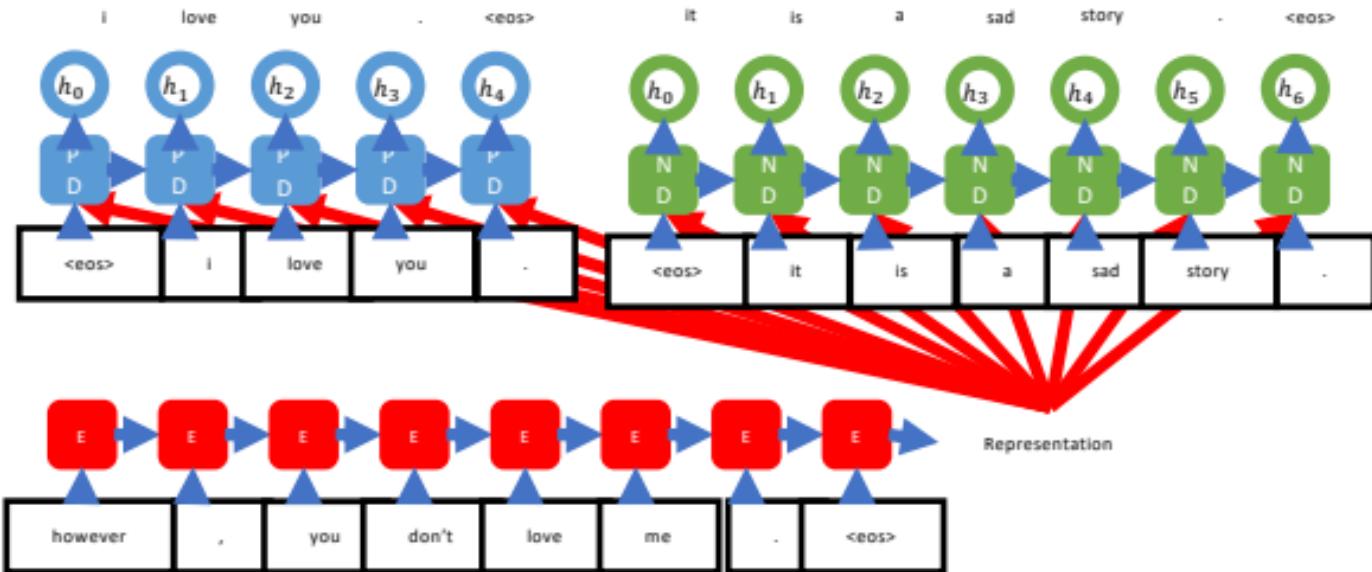
# Skip-thought (Kiros et al., NIPS2015)

- Given a sentence tuple
  - i love you.
  - however, you don't love me.
  - it is a sad story.
- Detailed encoding schemes
  - Uni-skip/ Bi-skip/ Combine-skip

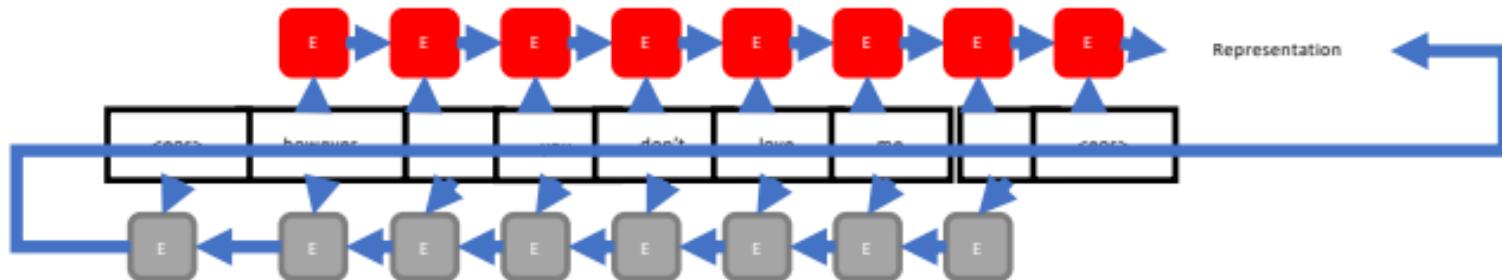
# Skip-thought (Kiros et al., NIPS2015)

- Given a sentence tuple
  - i love you.
  - however, you don't love me.
  - it is a sad story.
- Detailed encoding schemes
  - Uni-skip/ Bi-skip/ Combine-skip

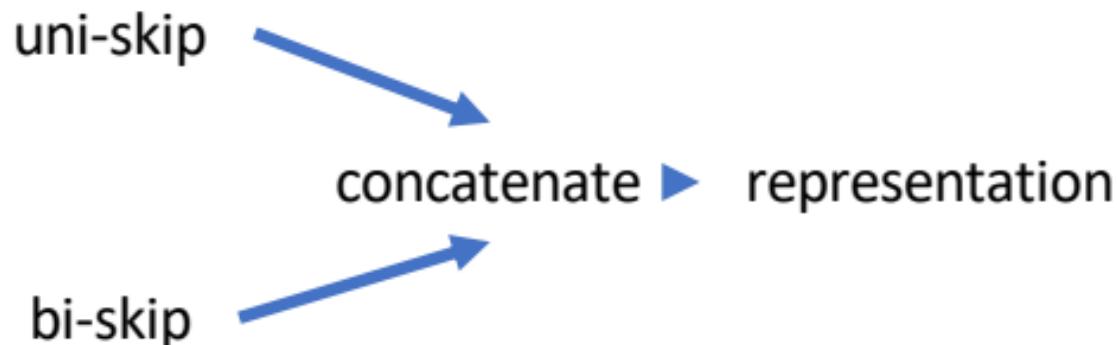
# Uni-Skip



# Bi-Skip



## Combine-Skip



# Trimming and Improving Skip-thought Vectors

- Skip-thought (Kiros et al., NIPS2015)
- Our hypotheses to improve skip-thought
  - **Neighborhood hypothesis**
  - Average+Max Connection
  - Word Vectors Initialization
- Comparison between our trimmed skip-thought model and the skip-thought model
- Conclusion

# Neighborhood Hypothesis

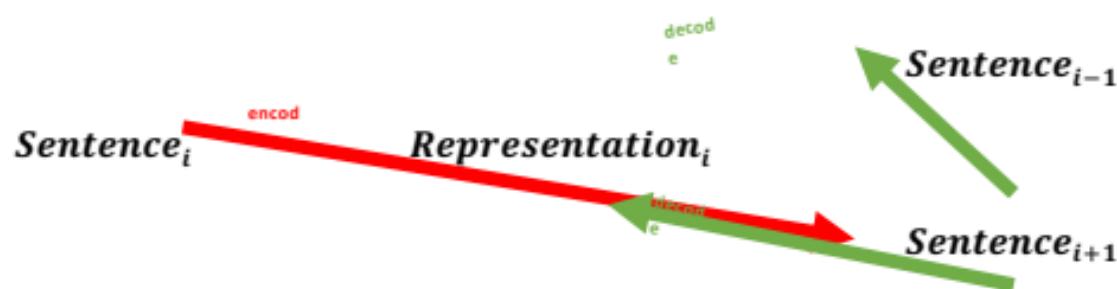
- Does this model really need a **previous** decoder and a **next** decoder?
- Hypothesis: Given the current sentence, inferring the **previous** sentence and inferring the **next** sentence both provide **same** supervision power.

# Neighborhood Hypothesis

- Skip-thought model



- Neighborhood Hypothesis

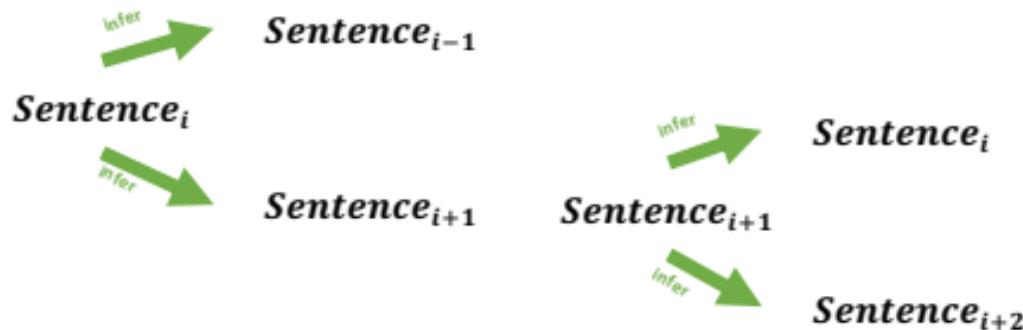


# Neighborhood Hypothesis

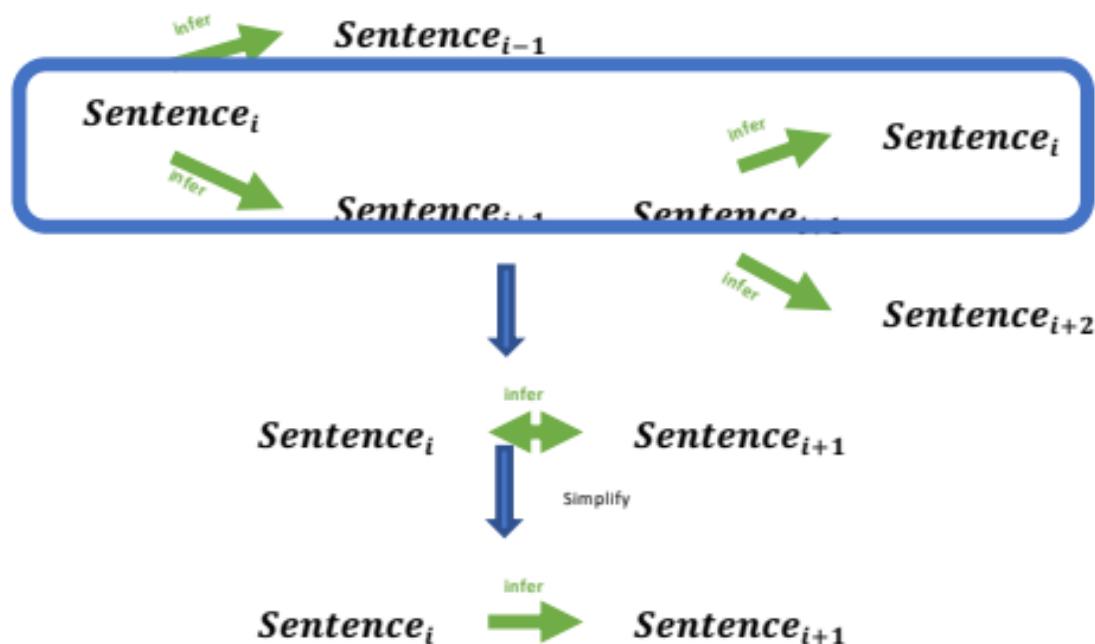
- Can we further simplify the skip-thought model?

**Yes!**

# Neighborhood Hypothesis

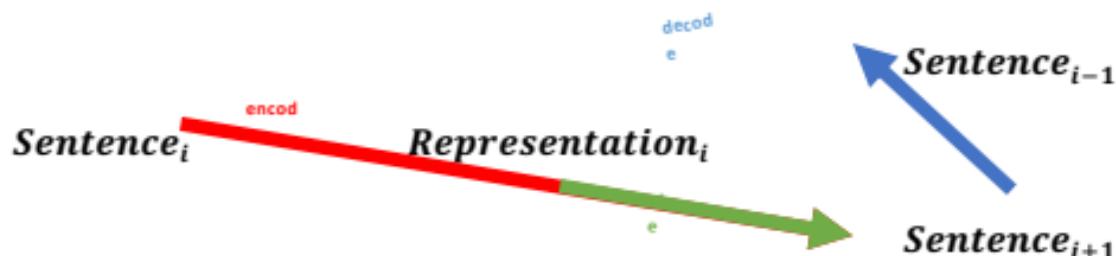


# Neighborhood Hypothesis



# Neighborhood Hypothesis

- Skip-thought Model



- Our **Trimmed** Skip-thought Model



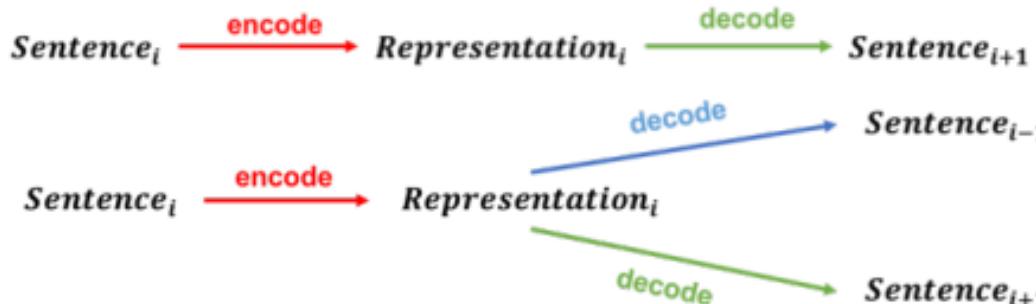
# Neighborhood Hypothesis

- **BookCorpus** dataset (Zhu et al., ICCV2015)
  - 74 million contiguous sentences from 7,000 books

## Encoder - Decoder

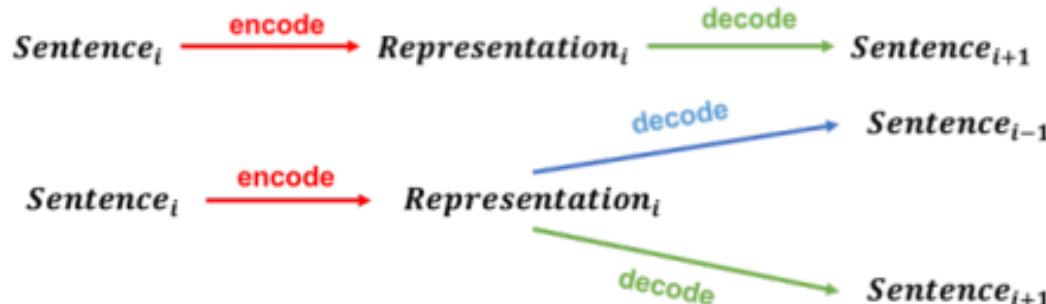
- Then, the sentence **encoder** was evaluated on 7 natural language processing (NLP) tasks.

# Neighborhood Hypothesis



Model	WE	Relationship of sentence pair			Classification on single sentence					
		$r$	$\rho$	MSE	MSRP (Acc/F1)	MR	CR	SUBJ	MPQA	TREC
<b>Plain Connection</b>										
bi-T-skip		0.8408	0.7649	0.2994	75.3 / <b>83.0</b>	76.1	80.3	92.3	87.5	86.6
uni-T-skip	word2vec	0.8349	0.7629	0.3084	73.7 / 81.9	75.7	82.1	91.3	87.4	86.4
C-T-skip		<b>0.8518</b>	<b>0.7808</b>	<b>0.2802</b>	<b>75.7 / 83.0</b>	76.8	<b>83.2</b>	<b>92.8</b>	<b>88.4</b>	87.5
bi-skip		0.8385	0.7618	0.3028	73.9 / 82.0	75.7	81.4	92.1	87.2	88.4
uni-skip	word2vec	0.8344	0.7586	0.3098	73.6 / 81.6	76.2	81.8	92.2	87.6	87.0
C-skip		0.8492	0.7738	0.2844	74.6 / 82.3	<b>77.0</b>	83.0	92.7	87.9	<b>89.2</b>

# Neighborhood Hypothesis



### Relationship of sentence pair

## Classification on single sentence

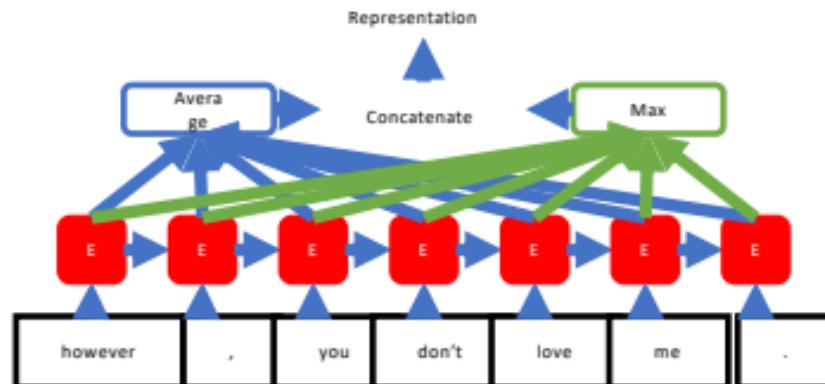
Model	WE	SICK			MSRP (Acc/F1)	MR	CR	SUBJ	MPQA	TREC
		r	$\rho$	MSE						
<b>Plain Connection</b>										
bi-T-skip	word2vec	0.8408	0.7649	0.2994	75.3 / 83.0	76.1	80.3	92.3	87.5	86.6
uni-T-skip		0.8349	0.7629	0.3084	73.7 / 81.9	75.7	82.1	91.3	87.4	86.4
C-T-skip		<b>0.8518</b>	<b>0.7808</b>	<b>0.2802</b>	<b>75.7 / 83.0</b>	<b>76.8</b>	<b>83.2</b>	<b>92.8</b>	<b>88.4</b>	<b>87.5</b>
bi-skip	word2vec	0.8385	0.7618	0.3028	73.9 / 82.0	75.7	81.4	92.1	87.2	88.4
uni-skip		0.8344	0.7586	0.3098	73.6 / 81.6	76.2	81.8	92.2	87.6	87.0
C-skip		0.8492	0.7738	0.2844	74.6 / 82.3	<b>77.0</b>	83.0	92.7	87.9	<b>89.2</b>

# Trimming and Improving Skip-thought Vectors

- Skip-thought
- Our hypotheses to improve skip-thought
  - Neighborhood hypothesis
  - **Average+Max Connection**
  - Word Vector Initialization
- Comparison between our trimmed skip-thought model and the skip-thought model
- Conclusion

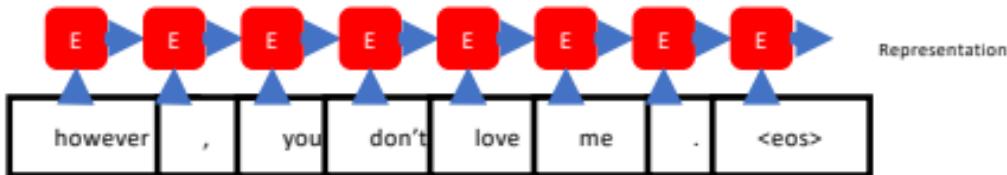
# Average+Max Connection

- Plain Connection (Kiros et al., NIPS2015)
- Average+Max Connection (Chen et al., arXiv2017)

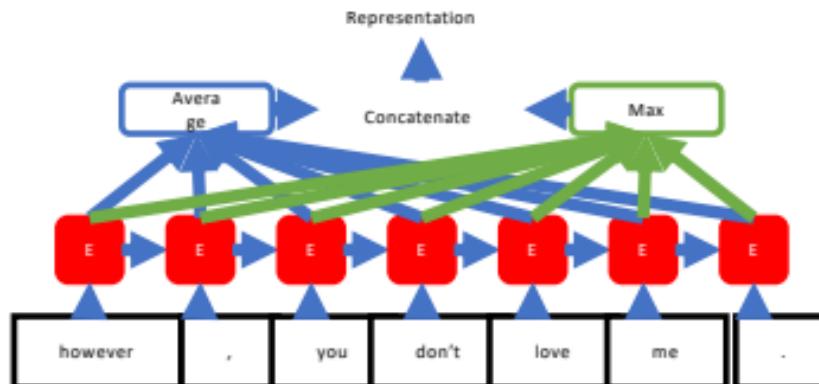


# Average+Max Connection

- Plain Connection (Kiros et al., NIPS2015)



- Average+Max Connection (Chen et al., arXiv2017)



# Average+Max Connection

- The average+max connection was proposed for **supervised** tasks, and specifically for Stanford Natural Language Inference corpus.
  - (Chen et al., arXiv2017; Bowman et al., EMNLP2015)



- We hypothesize that, our trimmed skip-thought model could also benefit from the average+max connection .

# Average+Max Connection

- The average+max connection was proposed for **supervised** tasks, and specifically for Stanford Natural Language Inference corpus.
  - (Chen et al., arXiv2017; Bowman et al., EMNLP2015)



- We hypothesize that, our trimmed skip-thought model could also **benefit** from the average+max connection .

# Average+Max Connection

Model	WE	Relationship of sentence pair			Classification on single sentence					
		SICK			MSRP (Acc/F1)	MR	CR	SUBJ	MPQA	TREC
		$r$	$\rho$	MSE						
<b>Plain Connection</b>										
bi-T-skip	word2vec	0.8408	0.7649	0.2994	75.3 / <b>83.0</b>	76.1	80.3	92.3	87.5	86.6
uni-T-skip		0.8349	0.7629	0.3084	73.7 / 81.9	75.7	82.1	91.3	87.4	86.4
C-T-skip		0.8518	0.7808	0.2802	<b>75.7 / 83.0</b>	<b>76.8</b>	<b>83.2</b>	<b>92.8</b>	<b>88.4</b>	87.5
<b>Average+Max Connection</b>										
bi-T-skip	word2vec	0.8463	0.7744	0.2894	73.3 / 81.6	74.4	78.6	91.3	86.2	88.8
uni-T-skip		0.8466	0.7705	0.2884	74.0 / 81.7	73.0	78.6	91.3	85.2	88.4
C-T-skip		<b>0.8598</b>	<b>0.7892</b>	<b>0.2654</b>	75.0 / 82.2	75.1	80.0	92.2	87.2	<b>90.0</b>

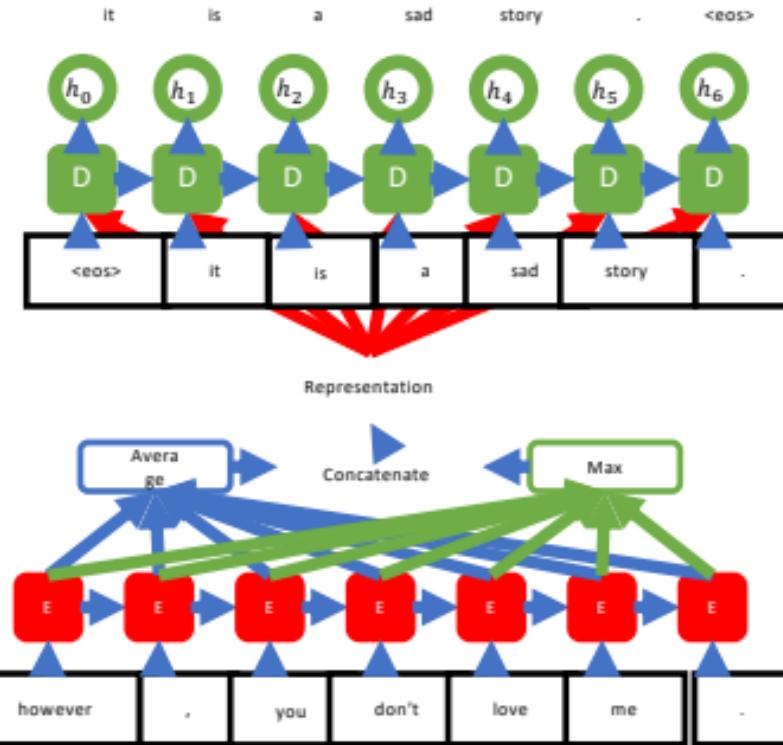
# Average+Max Connection

Model	WE	Relationship of sentence pair			Classification on single sentence					
		SICK			MSRP (Acc/F1)	MR	CR	SUBJ	MPQA	
		$r$	$\rho$	MSE						
<b>Plain Connection</b>										
bi-T-skip	word2vec	0.8408	0.7649	0.2994	75.3 / <b>83.0</b>	76.1	80.3	92.3	87.5	86.6
uni-T-skip		0.8349	0.7629	0.3084	73.7 / 81.9	75.7	82.1	91.3	87.4	86.4
C-T-skip		0.8518	0.7808	0.2802	<b>75.7 / 83.0</b>	<b>76.8</b>	<b>83.2</b>	<b>92.8</b>	<b>88.4</b>	87.5
<b>Average+Max Connection</b>										
bi-T-skip	word2vec	0.8463	0.7744	0.2894	73.3 / 81.6	74.4	78.6	91.3	86.2	88.8
uni-T-skip		0.8466	0.7705	0.2884	74.0 / 81.7	73.0	78.6	91.3	85.2	88.4
C-T-skip		<b>0.8598</b>	<b>0.7892</b>	<b>0.2654</b>	75.0 / 82.2	75.1	80.0	92.2	87.2	<b>90.0</b>

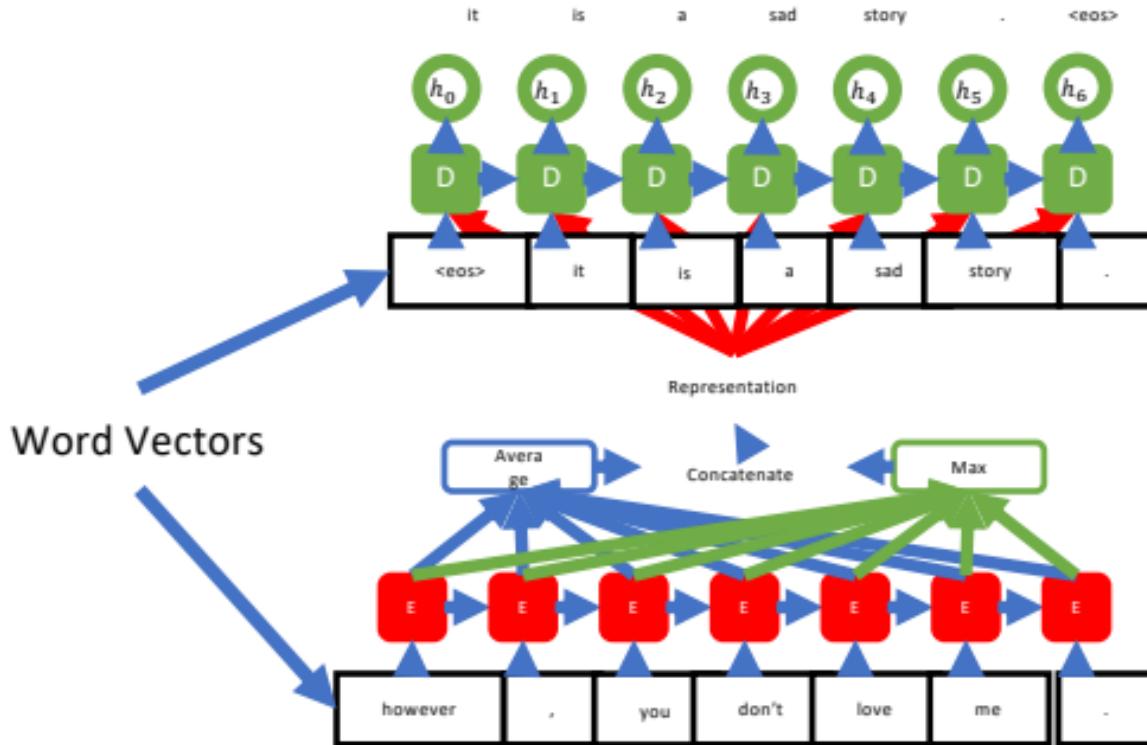
# Trimming and Improving Skip-thought Vectors

- Skip-thought
- Our hypotheses to improve skip-thought
  - Neighborhood hypothesis
  - Average+Max Connection
  - **Word Vectors Initialization**
- Comparison between our trimmed skip-thought model and the skip-thought model
- Conclusion

# Word Vectors Initialization



# Word Vectors Initialization



# Word Vectors Initialization

- Pretrained word vectors usually help the deep learning models to perform better on **supervised** tasks. (Collobert et al., JMLR2011)
  - word2vec (Mikolov et al., NIPS2013)
  - GloVe (Pennington et al., EMNLP2014)
- We hypothesize that, initializing our model with **pretrained word vectors** could also help the model to learn better sentence representation.

# Word Vectors Initialization

- Pretrained word vectors usually help the deep learning models to perform better on **supervised** tasks. (Collobert et al., JMLR2011)
  - word2vec (Mikolov et al., NIPS2013)
  - GloVe (Pennington et al., EMNLP2014)
- We hypothesize that, initializing our model with **pretrained word vectors** could also help the model to learn better sentence representation.

# Word Vectors Initialization

- Based on our trimmed skip-thought model with Average+Max connection, we implement models with 3 initializations.

Model	WE	SICK			MSRP (Acc/F1)	MR	CR	SUBJ	MPQA	TREC
		r	$\rho$	MSE						
<b>Average+Max Connection</b>										
bi-T-skip	random	0.8336	0.7612	0.3112	73.2 / 81.3	69.7	76.0	89.6	83.5	86.6
uni-T-skip		0.8293	0.7555	0.3180	72.5 / 81.0	67.3	74.9	89.0	81.1	83.6
C-T-skip		0.8458	0.7755	0.2902	74.7 / 82.1	70.4	76.7	90.4	83.8	84.8
bi-T-skip	GloVe	0.8444	0.7739	0.2922	75.1 / 82.4	74.4	79.5	90.9	85.3	87.6
uni-T-skip		0.8485	0.7711	0.2854	73.7 / 81.8	74.6	78.8	91.1	86.2	87.0
C-T-skip		0.8596	<b>0.7903</b>	0.2665	<b>75.4 / 82.6</b>	<b>75.6</b>	<b>80.4</b>	91.9	87.0	89.0
bi-T-skip	word2vec	0.8463	0.7744	0.2894	73.3 / 81.6	74.4	78.6	91.3	86.2	88.8
uni-T-skip		0.8466	0.7705	0.2884	74.0 / 81.7	73.0	78.6	91.3	85.2	88.4
C-T-skip		<b>0.8598</b>	0.7892	<b>0.2654</b>	75.0 / 82.2	75.1	80.0	<b>92.2</b>	<b>87.2</b>	<b>90.0</b>

# Word Vectors Initialization

Model	WE	Relationship of sentence pair			Classification on single sentence					
		$r$	$\rho$	MSE	MSRP (Acc/F1)	MR	CR	SUBJ	MPQA	TREC
<b>Average+Max Connection</b>										
bi-T-skip	random	0.8336	0.7612	0.3112	73.2 / 81.3	69.7	76.0	89.6	83.5	86.6
uni-T-skip		0.8293	0.7555	0.3180	72.5 / 81.0	67.3	74.9	89.0	81.1	83.6
C-T-skip		0.8458	0.7755	0.2902	74.7 / 82.1	70.4	76.7	90.4	83.8	84.8
bi-T-skip	GloVe	0.8444	0.7739	0.2922	75.1 / 82.4	74.4	79.5	90.9	85.3	87.6
uni-T-skip		0.8485	0.7711	0.2854	73.7 / 81.8	74.6	78.8	91.1	86.2	87.0
C-T-skip		0.8596	<b>0.7903</b>	0.2665	<b>75.4 / 82.6</b>	<b>75.6</b>	<b>80.4</b>	91.9	87.0	89.0
bi-T-skip	word2vec	0.8463	0.7744	0.2894	73.3 / 81.6	74.4	78.6	91.3	86.2	88.8
uni-T-skip		0.8466	0.7705	0.2884	74.0 / 81.7	73.0	78.6	91.3	85.2	88.4
C-T-skip		<b>0.8598</b>	0.7892	<b>0.2654</b>	75.0 / 82.2	75.1	80.0	<b>92.2</b>	<b>87.2</b>	<b>90.0</b>

## Furthermore...

- We wonder if adding more parameters could improve our model.

# Furthermore...

- Double-sized encoder gave us further improvement.

Model	WE	Relationship of sentence pair			Classification on single sentence						
		SICK			MSRP (Acc/F1)	MR	CR	SUBJ	MPQA	TREC	
		$r$	$\rho$	MSE							
<b>Doubled Encoder's Dimension vs. Results reported by [6]</b>											
Ours	bi-T-skip	word2vec	0.8503	0.7796	0.2823	74.4 / 82.2	74.8	80.3	91.8	87.0	88.2
	uni-T-skip		0.8486	0.7784	0.2857	74.3 / 82.4	72.9	78.0	90.7	85.7	86.4
	C-T-skip		<b>0.8611</b>	<b>0.7946</b>	<b>0.2634</b>	<b>74.5 / 82.2</b>	<b>75.4</b>	<b>80.3</b>	<b>92.2</b>	<b>87.4</b>	<b>88.4</b>
Kiros et al.	bi-skip [6]	random	0.8405	0.7696	0.2995	71.2 / 81.2	73.9	77.9	92.5	83.3	89.4
	uni-skip [6]		0.8477	0.7780	0.2872	73.0 / 81.9	75.5	79.3	92.1	86.9	91.4
	C-skip [6]		0.8584	0.7916	0.2687	73.0 / 82.0	<b>76.5</b>	80.1	<b>93.6</b>	87.1	<b>92.2</b>

# Furthermore...

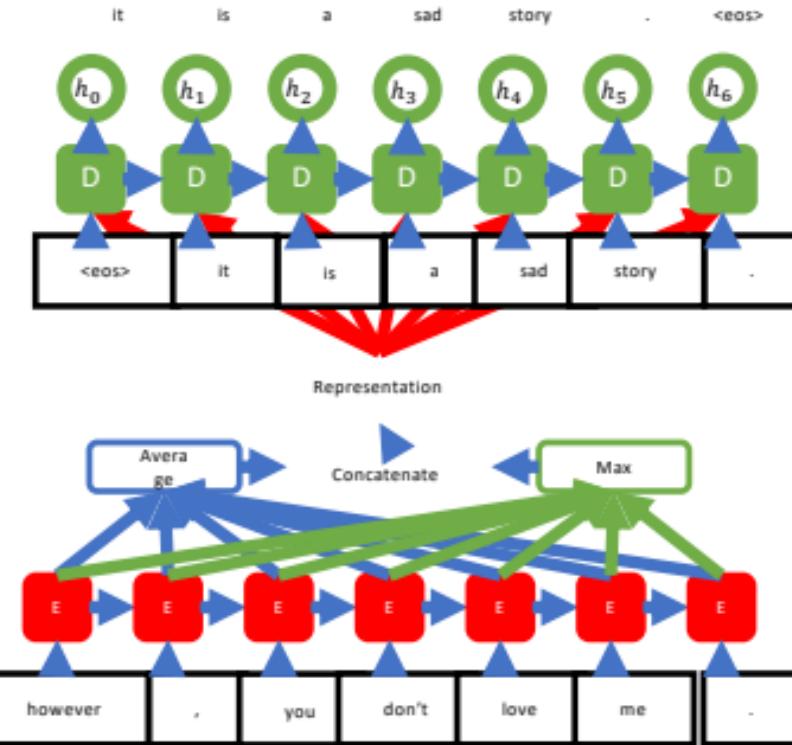
- Double-sized encoder gave us further improvement.

		Relationship of sentence pair			Classification on single sentence						
Model	WE	SICK			MSRP (Acc/F1)	MR	CR	SUBJ	MPQA	TREC	
		$r$	$\rho$	MSE							
<b>Doubled Encoder's Dimension vs. Results reported by [6]</b>											
Ours	bi-T-skip		0.8503	0.7796	0.2823	74.4 / 82.2	74.8	80.3	91.8	87.0	88.2
	uni-T-skip	word2vec	0.8486	0.7784	0.2857	74.3 / 82.4	72.9	78.0	90.7	85.7	86.4
	C-T-skip		<b>0.8611</b>	<b>0.7946</b>	<b>0.2634</b>	<b>74.5 / 82.2</b>	<b>75.4</b>	<b>80.3</b>	<b>92.2</b>	<b>87.4</b>	<b>88.4</b>
Kiros et al.	bi-skip [6]		0.8405	0.7696	0.2995	71.2 / 81.2	73.9	77.9	92.5	83.3	89.4
	uni-skip [6]	random	0.8477	0.7780	0.2872	73.0 / 81.9	75.5	79.3	92.1	86.9	91.4
	C-skip [6]		0.8584	0.7916	0.2687	73.0 / 82.0	<b>76.5</b>	80.1	<b>93.6</b>	87.1	<b>92.2</b>

# Trimming and Improving Skip-thought Vectors

- Skip-thought
- Our hypotheses to improve skip-thought
- **Comparison between our trimmed skip-thought model and the skip-thought model**
  - Number of Parameters
  - Training Time
- Conclusion

# Our Trimmed Skip-thought



# Our Trimmed Skip-thought

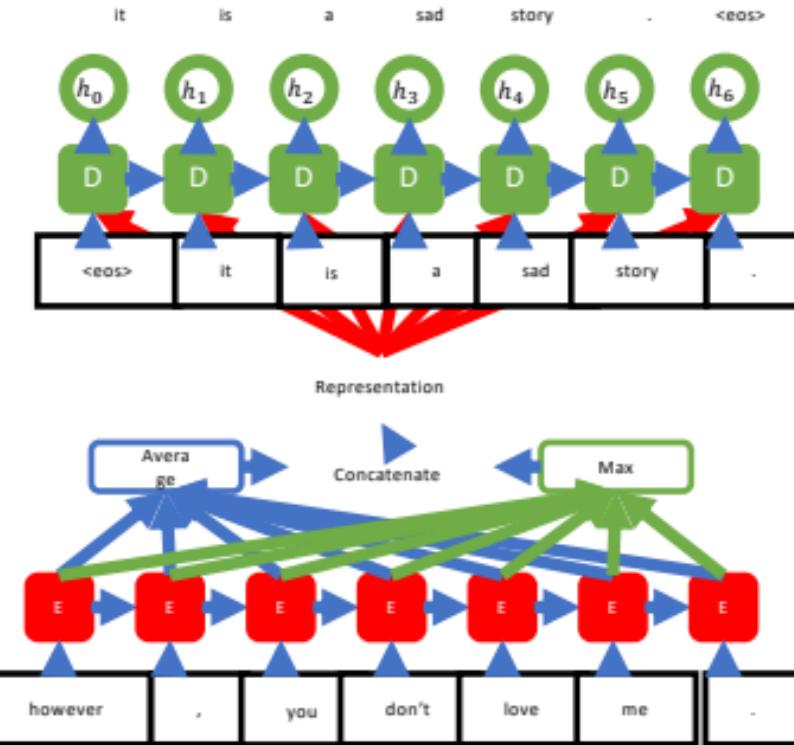
Skip-thought (Kiros et al., NIPS2016)	620*20000
Our Trimmed and Improved Skip-thought	300*20000



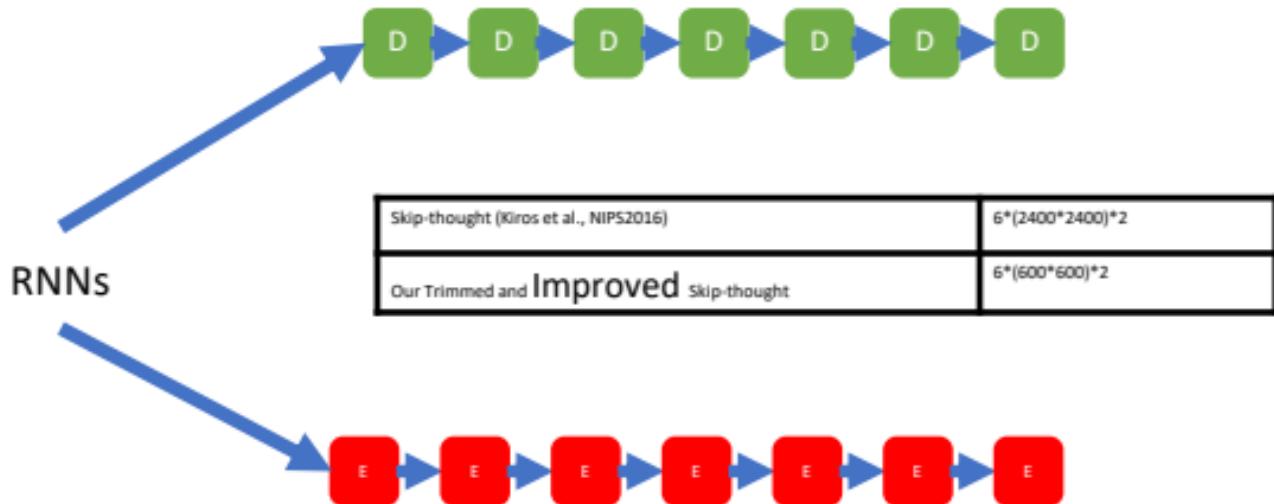
Word Vectors



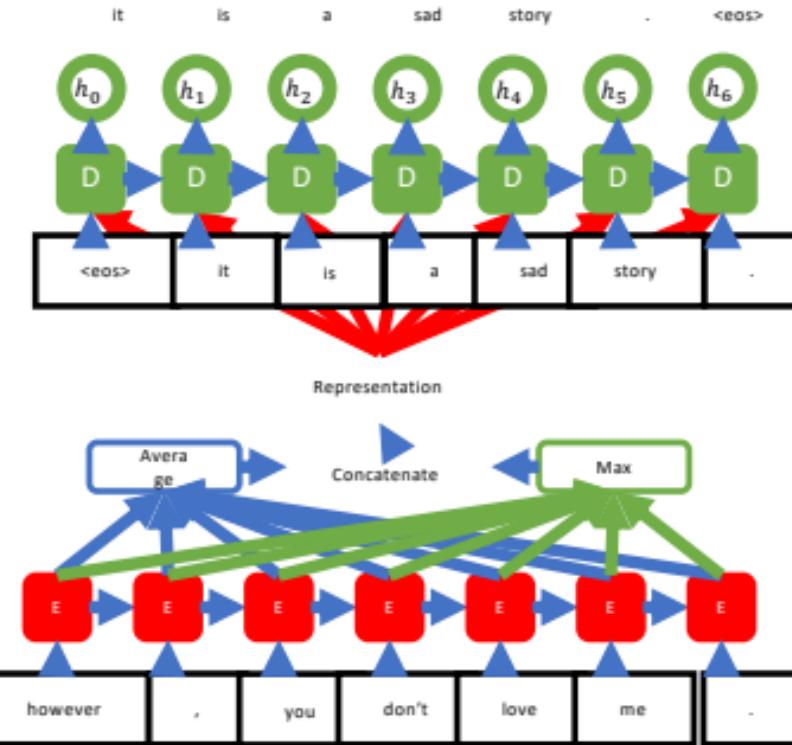
# Our Trimmed Skip-thought



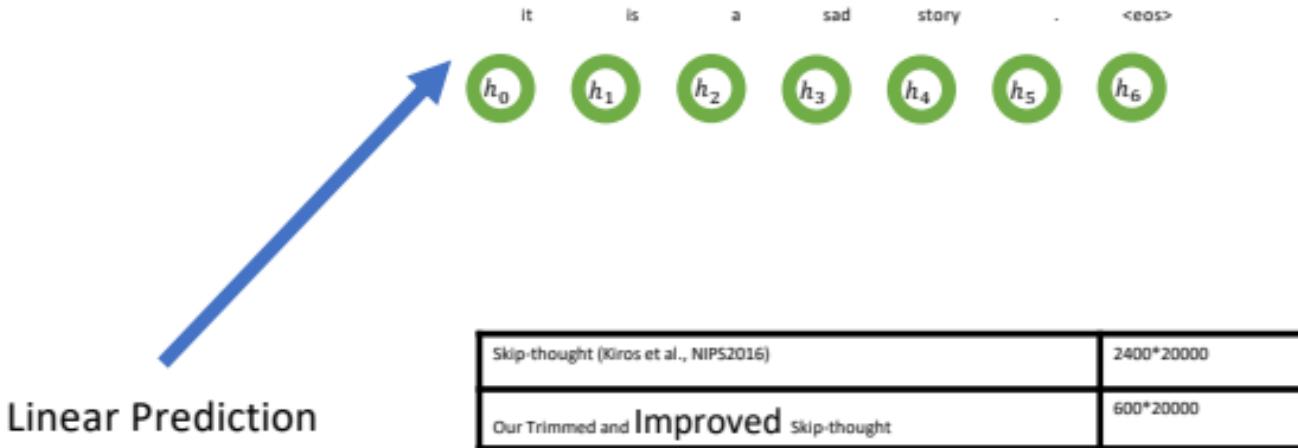
# Our Trimmed Skip-thought



# Our Trimmed Skip-thought



# Our Trimmed Skip-thought



# Number of Parameters

Model	RNNs	Word Vectors	Linear Prediction
uni-T-skip (ours)	4.32M		
bi-T-skip (ours)	3.24M		
uni-T-skip-double (ours)	10.80M	6M	12M
bi-T-skip-double (ours)	6.48M		
uni-skip (Kiros et al., NIPS2015)	69.12M		
bi-skip (Kiros et al., NIPS2015)	51.84M	12.4M	48M

"RNNs" refers to recurrent networks in the encoder and the decoder.

"Word Embedding" refers to all word vectors in unsupervised training.

"Linear Prediction" refers to the linear prediction layer in the decoder.

# Training Time

Model	Training Time
Skip-thought (Kiros et al., NIPS2015)	2 weeks

# Training Time

Model	Training Time
Skip-thought (Kiros et al., NIPS2015)	2 weeks
Skip-thought (our implementation)	4 days

# Training Time

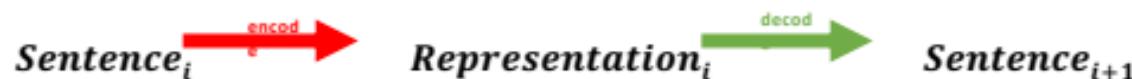
Model	Training Time
Skip-thought (Kiros et al., NIPS2015)	2 weeks
Skip-thought (our implementation)	4 days
Our Trimmed and Improved Skip-thought	1 day

# Trimming and Improving Skip-thought Vectors

- Skip-thought
- Our hypotheses to improve skip-thought
- Comparison between our trimmed skip-thought model and the skip-thought model
- **Conclusion**

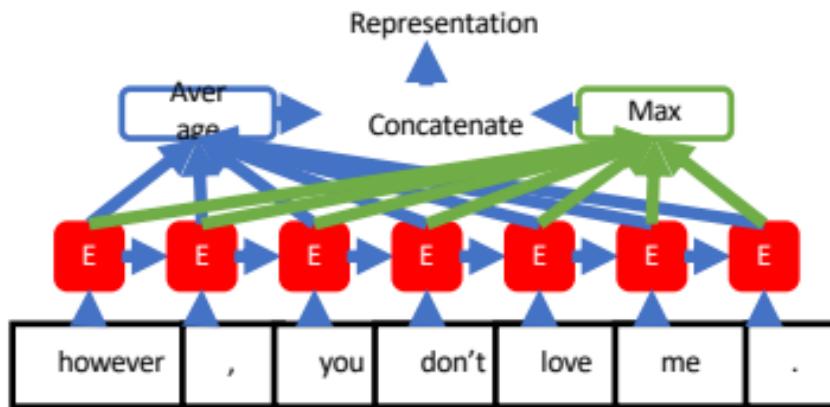
Conclusion: We improved skip-thought model

- by dropping one decoder



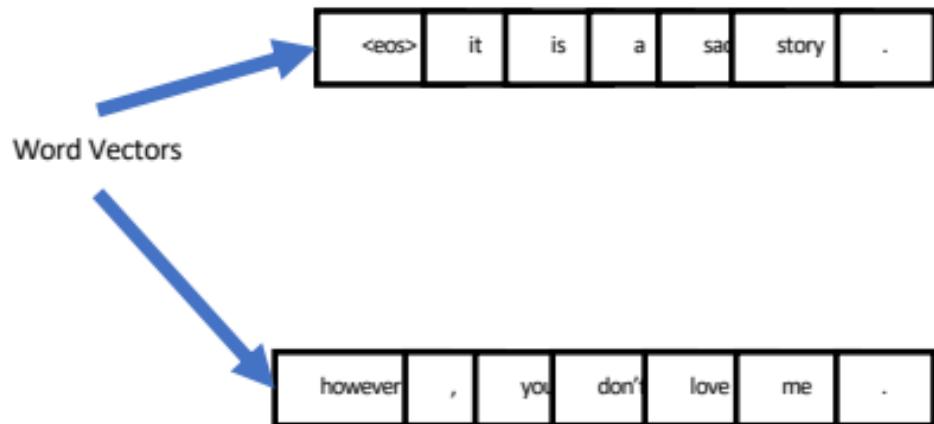
# Conclusion: We improved skip-thought model

- by applying the **average+max connection** between the encoder and the decoder



# Conclusion: We improved skip-thought model

- by **initializing** the model with **pretrained word vectors** instead of random values



## Conclusion: We improved skip-thought model

- by **accelerating** the training procedure, because we cut out 80% parameters in the skip-thought model.

Model	Training Time
Skip-thought (Kiros et al., NIPS2015)	2 weeks
Skip-thought (our implementation)	4 days
Our Trimmed and Improved Skip-thought	1 day

# Committee & Collaborators

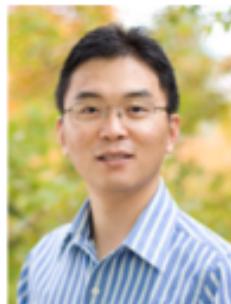
- Committee members

- Virginia R. de Sa, CogSci
- Benjamin K. Bergen, CogSci
- Jeffrey L. Elman, CogSci
- Julian J. McAuley, CSE



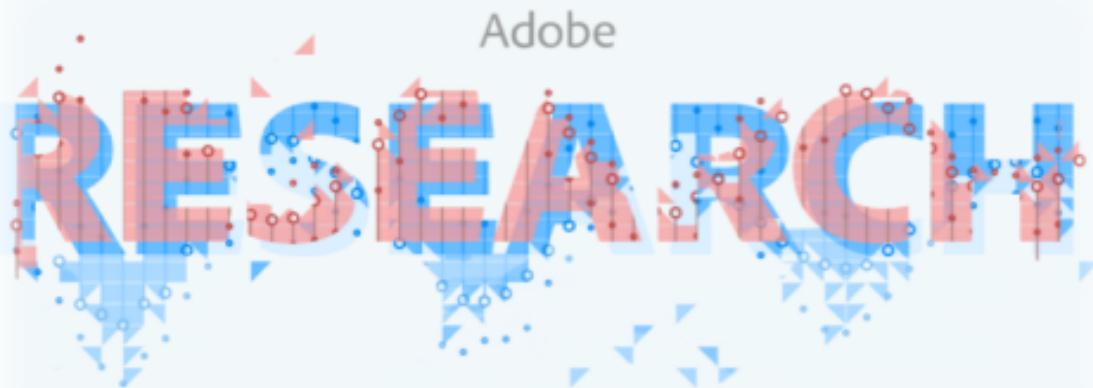
- Collaborators

- Hailin Jin, Chen Fang, Zhaowen Wang
- Researchers at Adobe research lab



# Acknowledgements

- Many thanks to Adobe Research Lab for GPUs support, also to NVIDIA for DGX-1 trial.



# Acknowledgements

- Many thanks to cohort, and Marta.



Q & A