Going Deeper with Images and Natural Language

Ph.D. Preliminary Examination

Yufeng Ma

February 16, 2018

Department of Computer Science
Virginia Tech, Blacksburg, VA 24061 USA

Advisory Committee:
Dr. Weiguo (Patrick) Fan
Dr. Edward A. Fox
Dr. G. Alan Wang
Dr. Bert Huang
Dr. Zhongju (John) Zhang

Overview

1. Introduction
   - Motivation
   - Problem Statement
   - Deep Learning Backgrounds
   - Hypotheses
   - Research Questions
2. Improved Image Captioning with Adversarial Loss
   - Model Architecture
   - Optimizations and Experiment Results
3. Congruence Measure between Image and Sentences
   - Pseudo Supervised Training
   - Loss Functions and Preliminary Results
4. Image Aspect Mining
   - Task Descriptions and Approaches
5. Conclusions and Research Timeline
6. Acknowledgements

Motivation

Motivation

Alan Turing

We may hope that machines will eventually compete with men in all purely intellectual fields.

Motivation

Alan Turing

We may hope that machines will eventually compete with men in all purely intellectual fields.

Current Progress in Artificial Intelligence

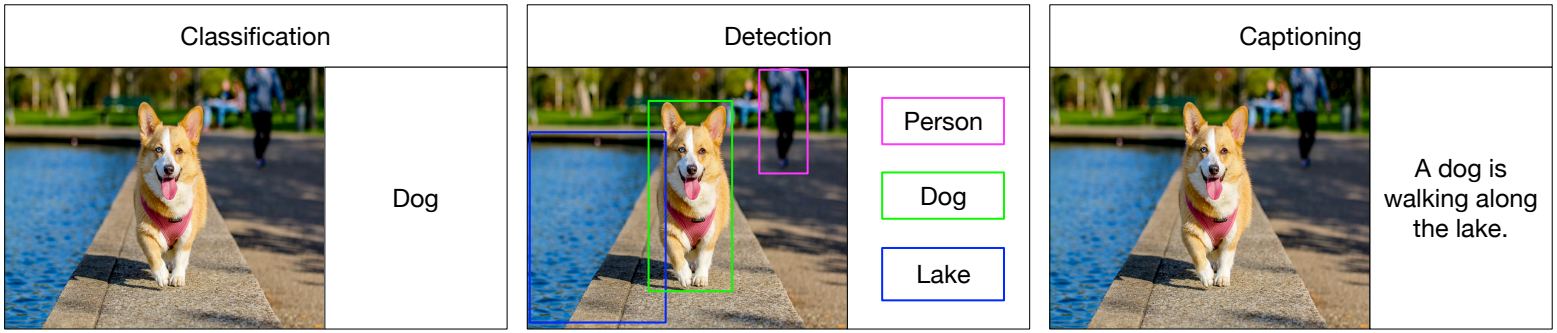
Motivation

Alan Turing

We may hope that machines will eventually compete with men in all purely intellectual fields.

Current Progress in Artificial Intelligence

Vision:



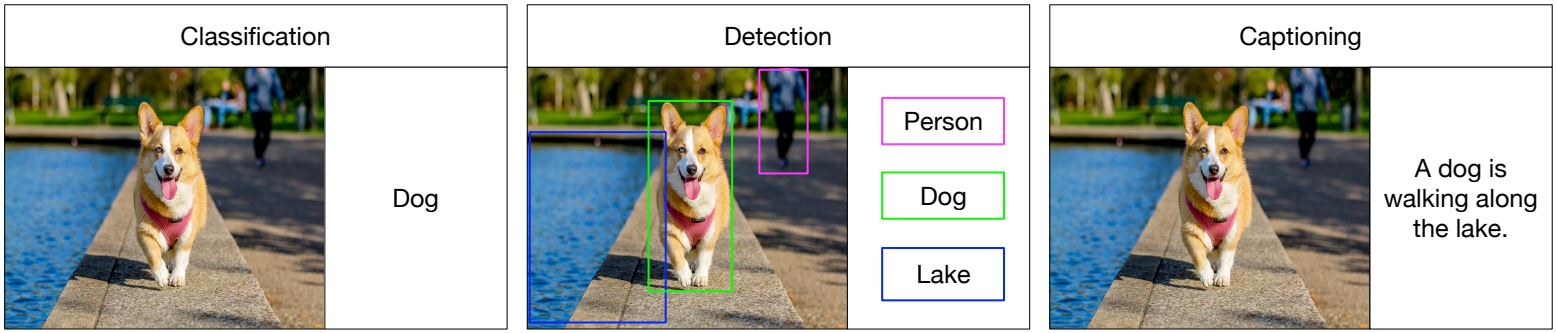
Motivation

Alan Turing

We may hope that machines will eventually compete with men in all purely intellectual fields.

Current Progress in Artificial Intelligence

Vision:



NLP: machine translation, text summarization, sentiment analysis, etc.

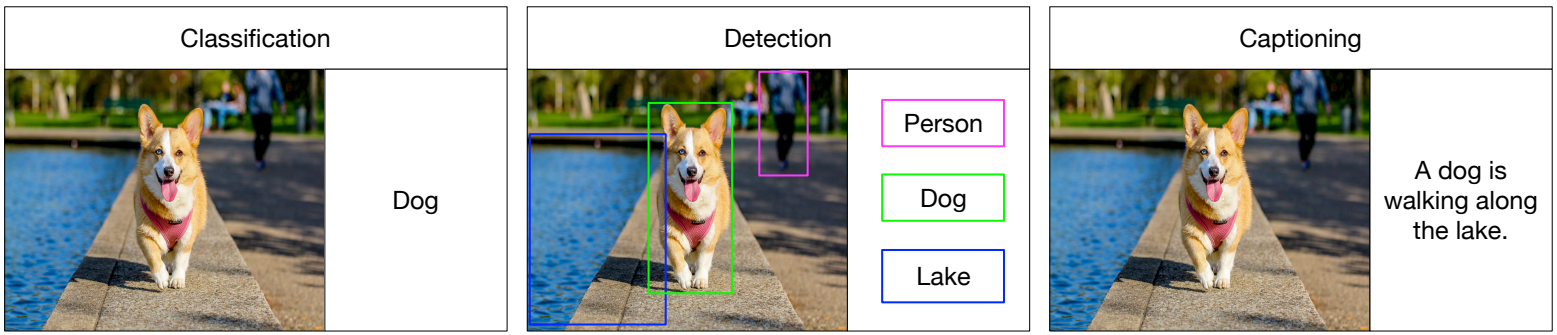
Motivation

Alan Turing

We may hope that machines will eventually compete with men in all purely intellectual fields.

Current Progress in Artificial Intelligence

Vision:



NLP: machine translation, text summarization, sentiment analysis, etc.

Encouraging Progress

- CNN: *image classification, object detection, segmentation, ...*
- RNN: *speech recognition, machine translation, stock price prediction, ...*

Motivation

Motivation

What's next? Image $\Leftrightarrow$ NLP

- Connecting semantics in text with entities in images.
- Fine-grained inference w.r.t. aspects in text and images.

Motivation

What's next? Image $\Leftrightarrow$ NLP

- Connecting semantics in text with entities in images.
- Fine-grained inference w.r.t. aspects in text and images.

Why?

- Huge amount of tweets/posts/reviews with both images and text.
- Humans are inherently able to connect their semantics effortlessly.
- It's unable for readers to literally look through all of them.

Problem Statement

Problem Statement

1. Generating high quality image descriptions
   that are indistinguishable to ones written by humans.
2. Relating a set of review sentences to images.
3. Recognizing aspects in review text,
   matching with regions in images, and inferring fine-grained ratings.

Deep Learning Backgrounds

Deep Learning Backgrounds

- Deep Supervised Learning
  - Objective Function
  - Regularization

Deep Learning Backgrounds

- Deep Supervised Learning
  - Objective Function
  - Regularization
- Deep Unsupervised Learning - GANs
  - Generator & Discriminator

Deep Learning Backgrounds

- Deep Supervised Learning
  - Objective Function
  - Regularization
- Deep Unsupervised Learning - GANs
  - Generator & Discriminator
- Optimization
  - Stochastic Gradient Descent
  - Momentum, Adagrad, RMSProp, Adam

Deep Learning Backgrounds

- Deep Supervised Learning
  - Objective Function
  - Regularization
- Deep Unsupervised Learning - GANs
  - Generator & Discriminator
- Optimization
  - Stochastic Gradient Descent
  - Momentum, Adagrad, RMSProp, Adam
- Backpropagation
  - Chain rule, Jacobian matrix

Deep Learning Backgrounds

- Deep Supervised Learning
  - Objective Function
  - Regularization
- Deep Unsupervised Learning - GANs
  - Generator & Discriminator
- Optimization
  - Stochastic Gradient Descent
  - Momentum, Adagrad, RMSProp, Adam
- Backpropagation
  - Chain rule, Jacobian matrix
- Neural Networks
  - Convolutional Neural Network
  - Recurrent Neural Network - LSTM

Deep Learning Backgrounds

- Deep Supervised Learning
  - Objective Function
  - Regularization
- Deep Unsupervised Learning - GANs
  - Generator & Discriminator
- Optimization
  - Stochastic Gradient Descent
  - Momentum, Adagrad, RMSProp, Adam
- Backpropagation
  - Chain rule, Jacobian matrix
- Neural Networks
  - Convolutional Neural Network
  - Recurrent Neural Network - LSTM
- Attention Mechanism

Hypotheses

Hypotheses

Principal Assumptions

- Neural networks can estimate connections between images and texts.
- CNN is able to extract abstractive concepts in images.
- RNN can accurately deliver textual semantics.

Hypotheses

Principal Assumptions

- Neural networks can estimate connections between images and texts.
- CNN is able to extract abstractive concepts in images.
- RNN can accurately deliver textual semantics.

1. Image caption quality can be improved by training with GANs.
2. Relevances between images and review sentences can be estimated with unsupervised learning.
3. Correlation between aspects in text and images can be computed; and fine-grained ratings can be accurately figured.

Research Questions

Research Questions

1. Improve Image Caption Qualities

How and by how much can we improve the quality of image captions generated by machines?

Research Questions

1. Improve Image Caption Qualities

How and by how much can we improve the quality of image captions generated by machines?

2. Match Review Sentences to Image

How can we measure the congruence of sentence-image pair if there is no human annotated label for model training?

Research Questions

1. Improve Image Caption Qualities

How and by how much can we improve the quality of image captions generated by machines?

2. Match Review Sentences to Image

How can we measure the congruence of sentence-image pair if there is no human annotated label for model training?

3. Image Aspect Mining

How can we connect aspects mined from text and ones detected in images and do fine-grained inference?

Overview

1. Introduction
   - Motivation
   - Problem Statement
   - Deep Learning Backgrounds
   - Hypotheses
   - Research Questions
2. Improved Image Captioning with Adversarial Loss
   - Model Architecture
   - Optimizations and Experiment Results
3. Congruence Measure between Image and Sentences
   - Pseudo Supervised Training
   - Loss Functions and Preliminary Results
4. Image Aspect Mining
   - Task Descriptions and Approaches
5. Conclusions and Research Timeline
6. Acknowledgements

Image Captioning

Image Captioning

Research Question

How and by how much can we improve the quality of image captions generated by machines?

Image Captioning

Research Question

How and by how much can we improve the quality of image captions generated by machines?

What's image captioning?

Image Captioning

Research Question

How and by how much can we improve the quality of image captions generated by machines?

What's image captioning?

Alternative Captions:

1. A picture of a dog laying on the ground
2. Dog snoozing by a bike on the edge of a cobblestone street
3. The white dog lays next to the bicycle on the sidewalk.
4. A white dog is sleeping on a street and a bicycle.
5. A puppy rests on the street next to a bicycle.

Image Captioning

Image Captioning

Impacts

- Help visually impaired perceive the surroundings.
- Captions alleviate the difficulty of image retrieval.
- Tweets/posts/reviews with images and text: better understanding.

Image Captioning

Impacts

- Help visually impaired perceive the surroundings.
- Captions alleviate the difficulty of image retrieval.
- Tweets/posts/reviews with images and text: better understanding.

Related Works

- Visual elements recognition $\rightarrow$ Language modeling (Farhadi et al. 2010, Kulkarni et al. 2013)
- Encoder (CNN) + Decoder (RNN) + Attention (Karpathy et al. 2015, Xu et al. 2015, Lu et al. 2017, Anderson et al. 2017)
- REINFORCE for performance boosting (Rennie et al. 2016)
- GANs for discrete data generation (Jang et al. 2016)

Image Captioning

Impacts

- Help visually impaired perceive the surroundings.
- Captions alleviate the difficulty of image retrieval.
- Tweets/posts/reviews with images and text: better understanding.

Related Works

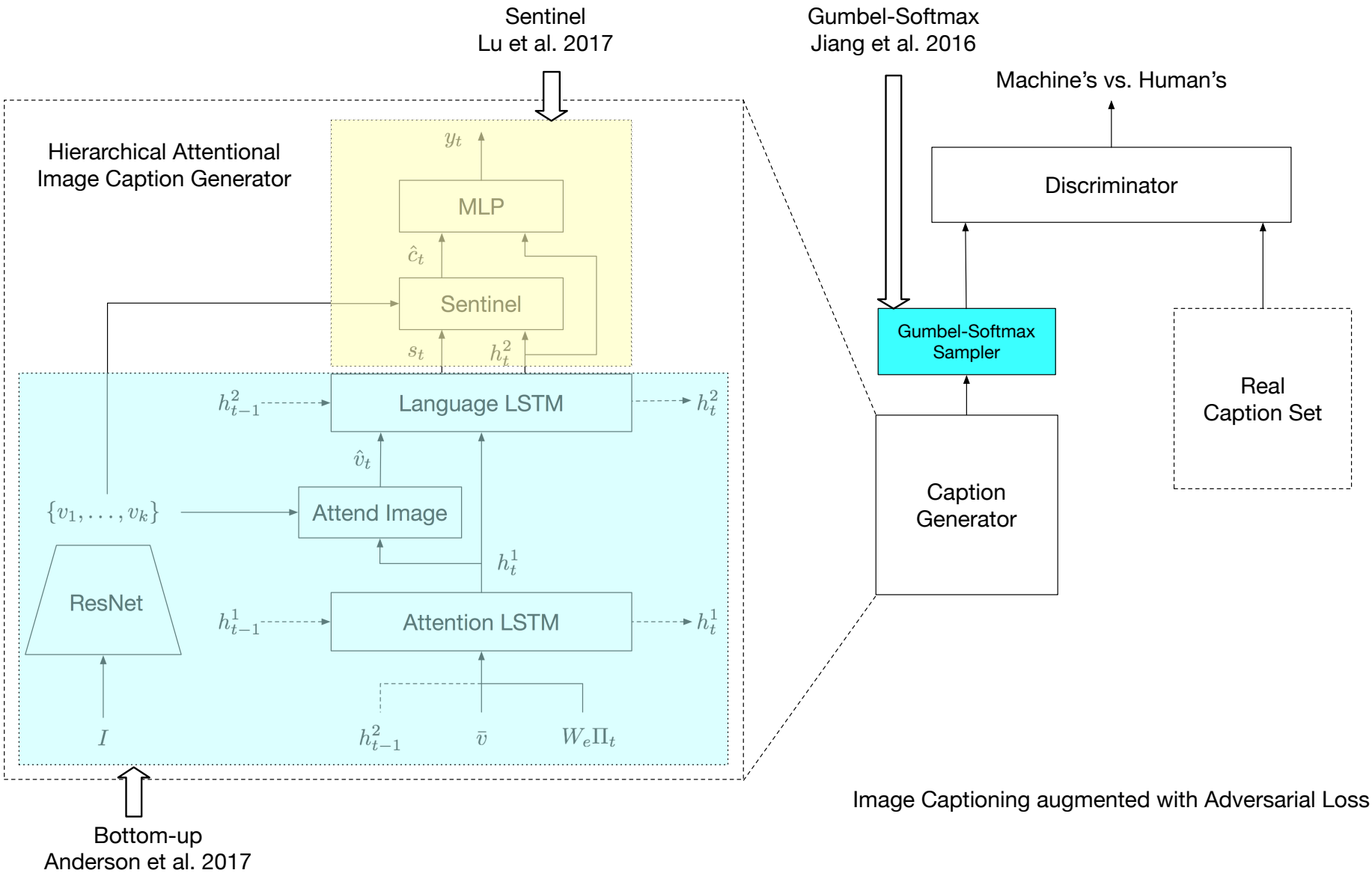
- Visual elements recognition $\rightarrow$ Language modeling (Farhadi et al. 2010, Kulkarni et al. 2013)
- Encoder (CNN) + Decoder (RNN) + Attention (Karpathy et al. 2015, Xu et al. 2015, Lu et al. 2017, Anderson et al. 2017)
- REINFORCE for performance boosting (Rennie et al. 2016)
- GANs for discrete data generation (Jang et al. 2016)

Issues and Challenges

Differences between machine generated and human written captions

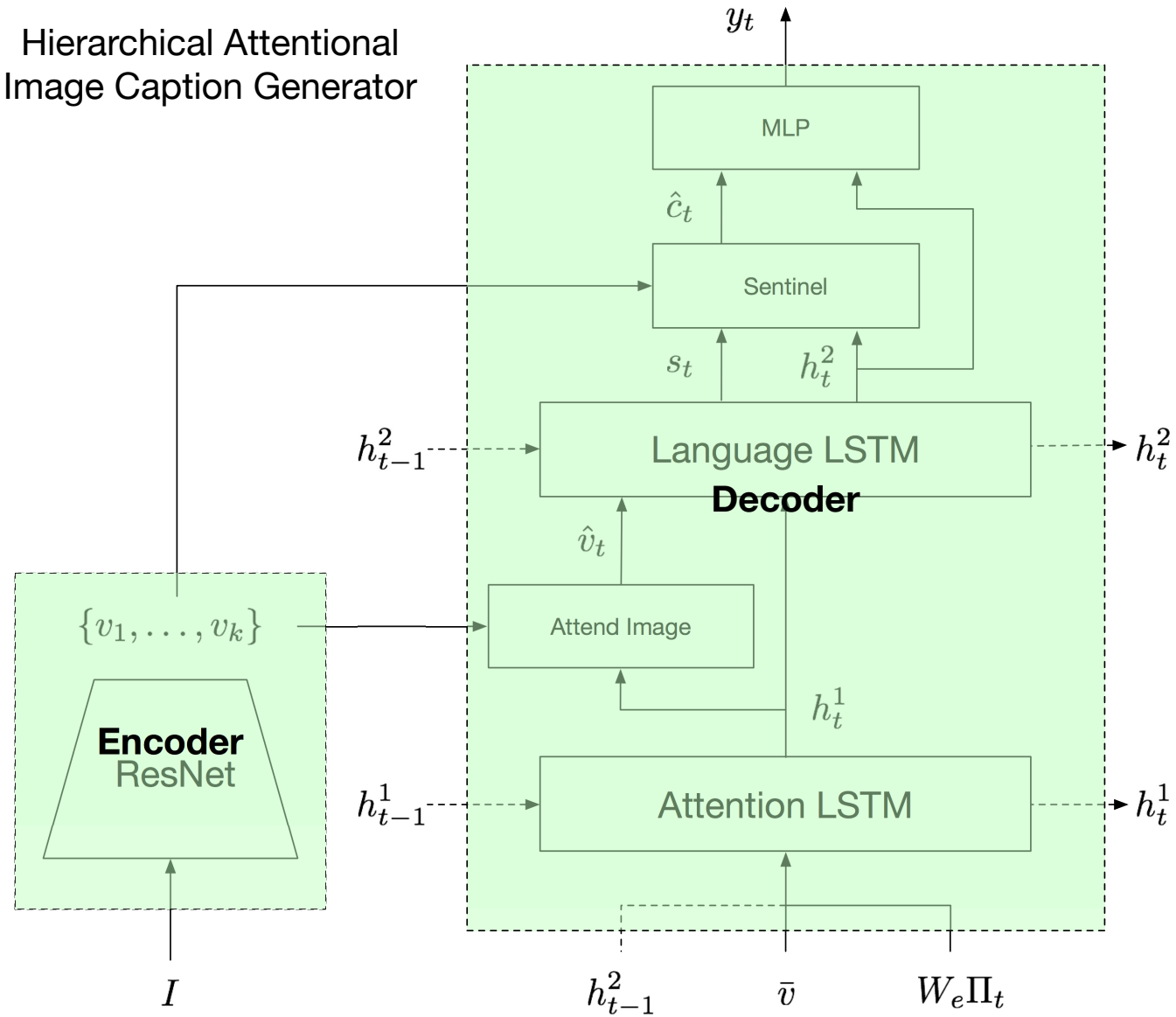
Model Architecture

Model Architecture

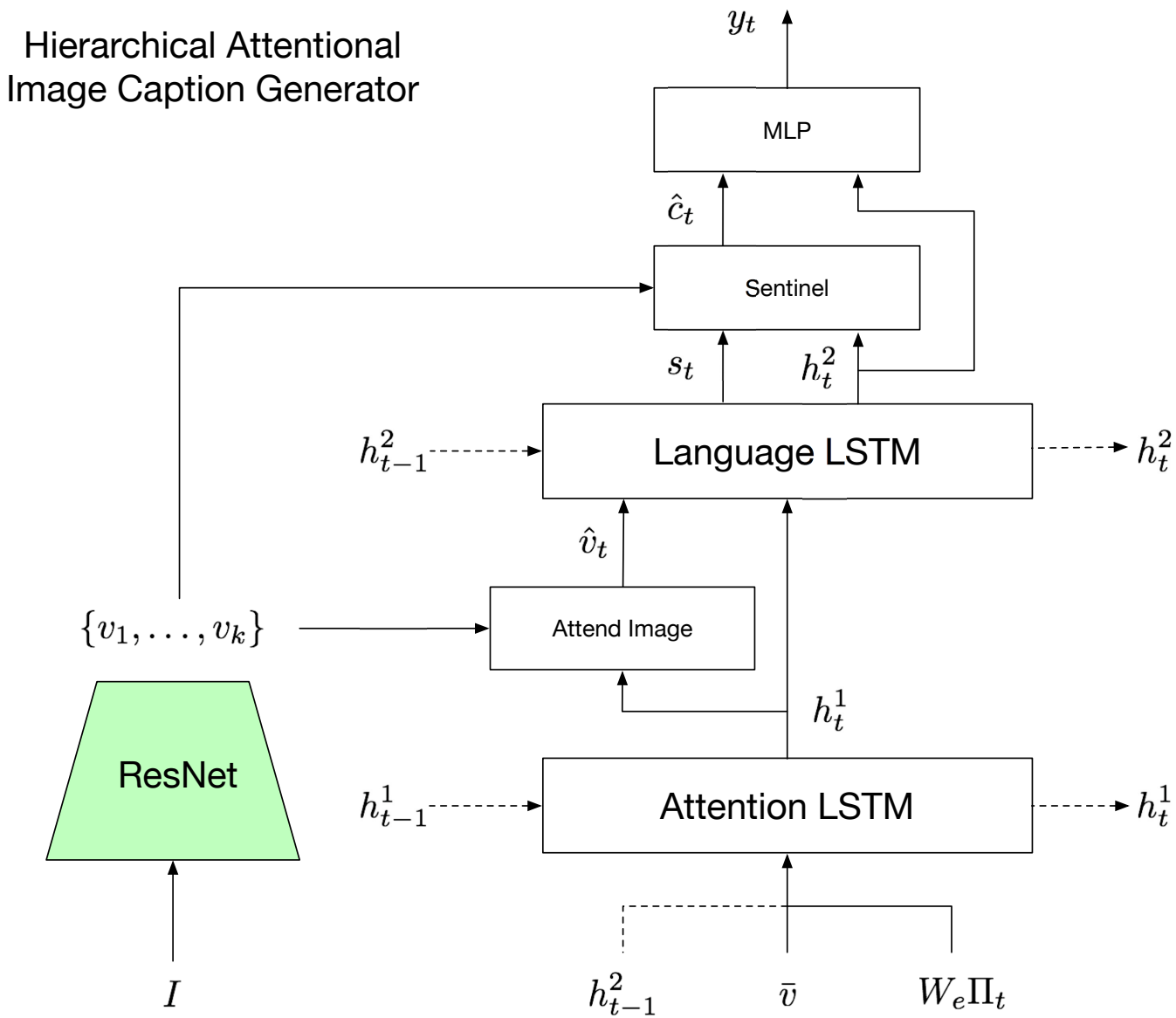


Model Architecture

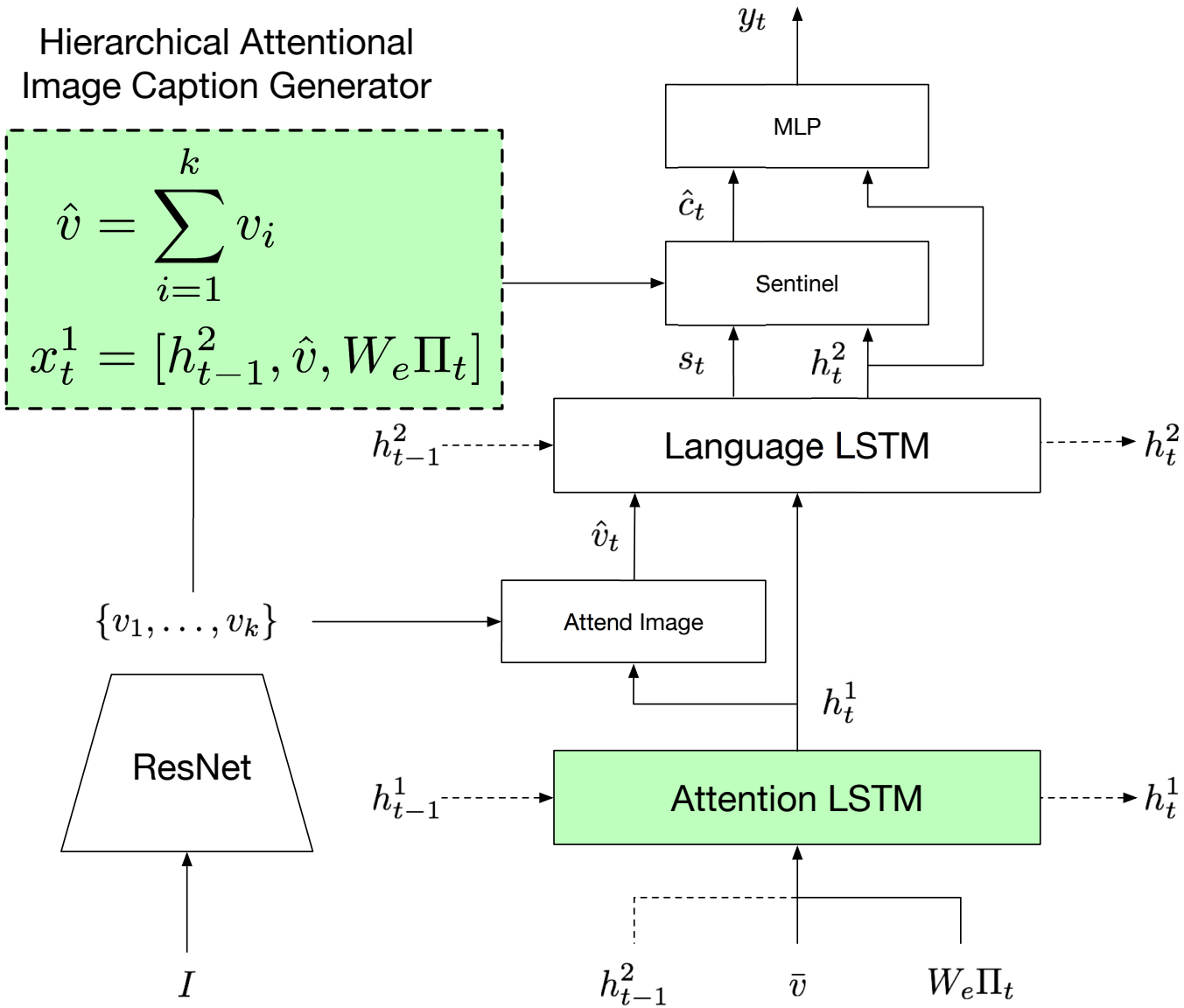
Model Architecture



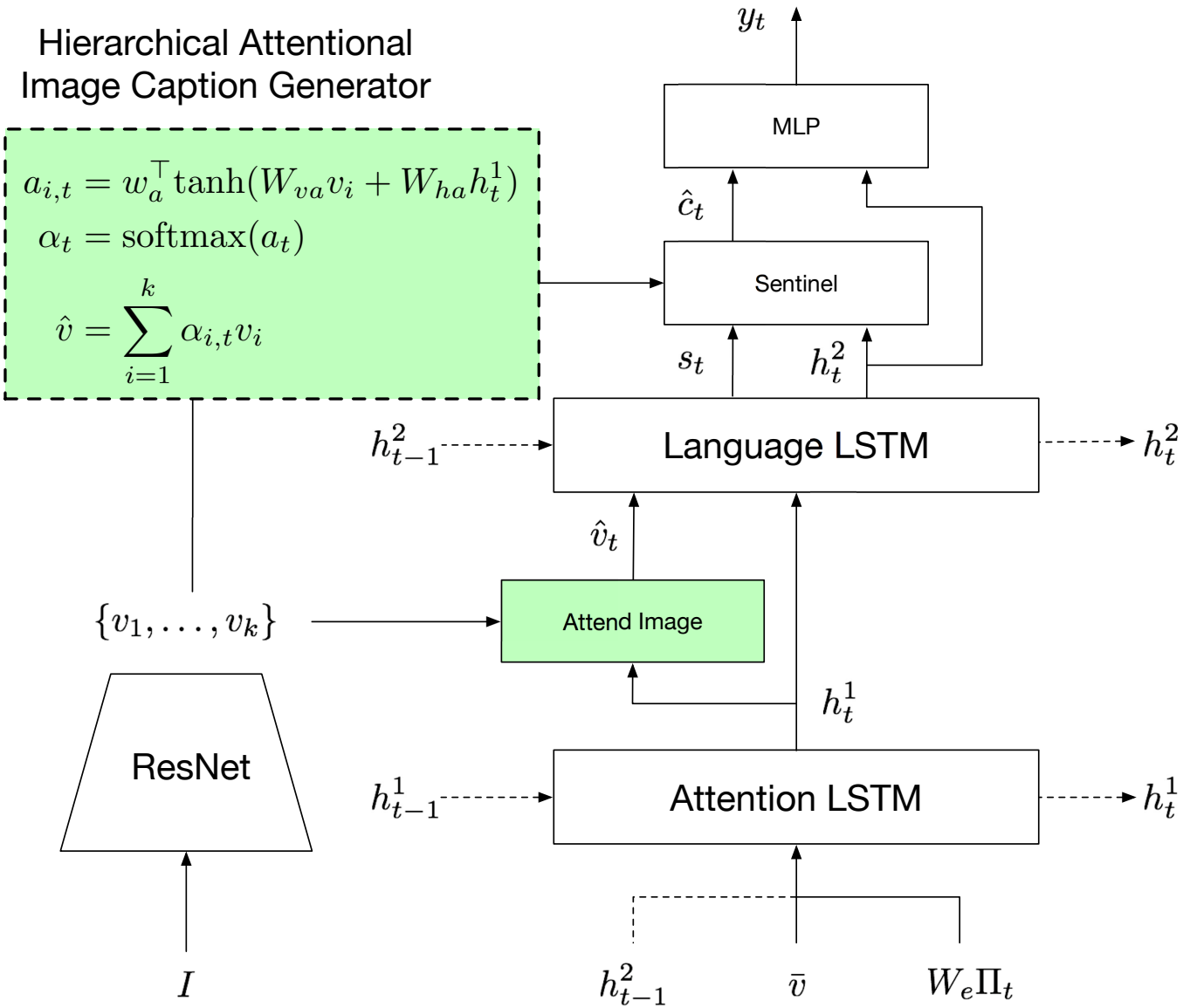
Model Architecture



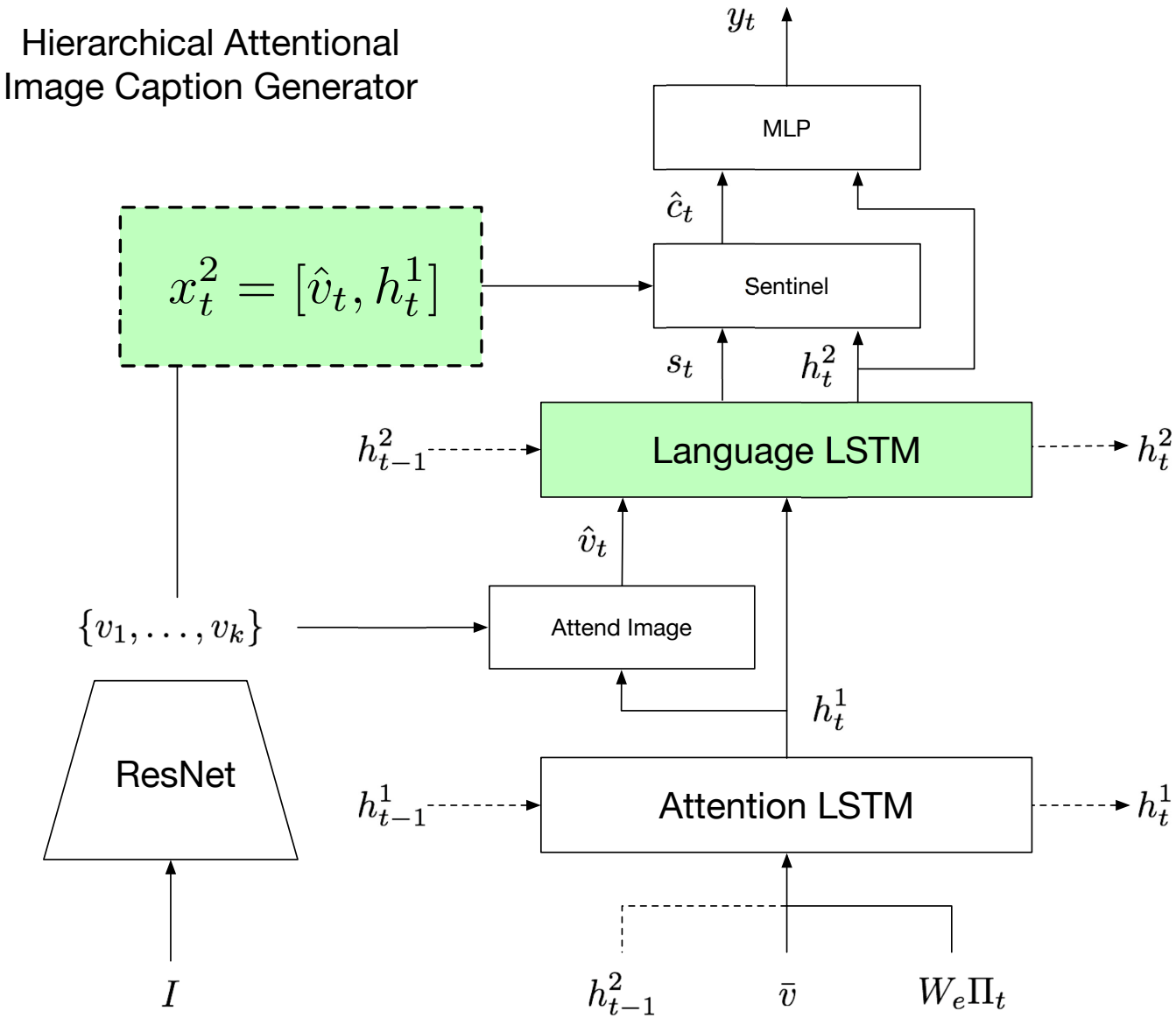
Model Architecture



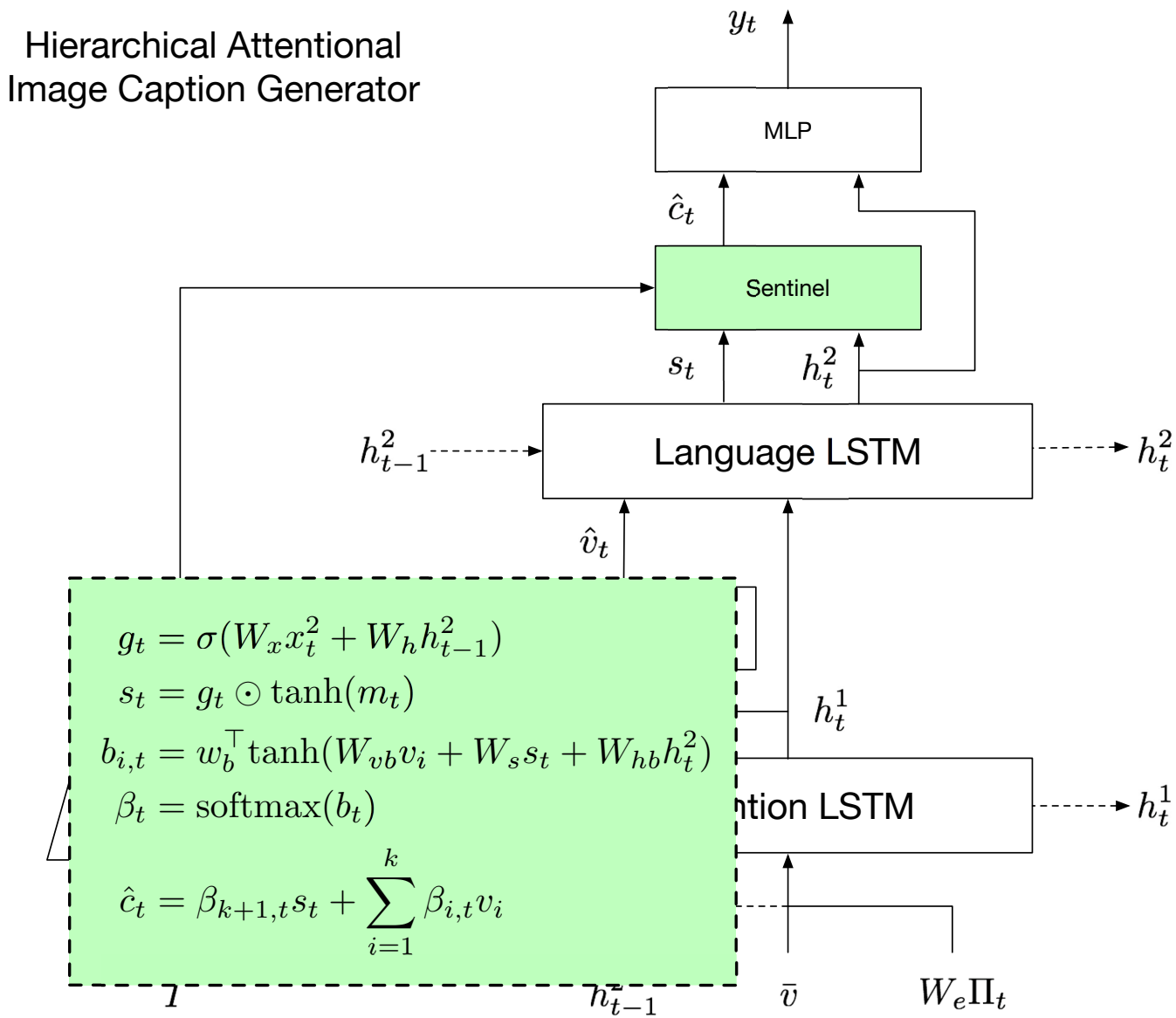
Model Architecture



Model Architecture

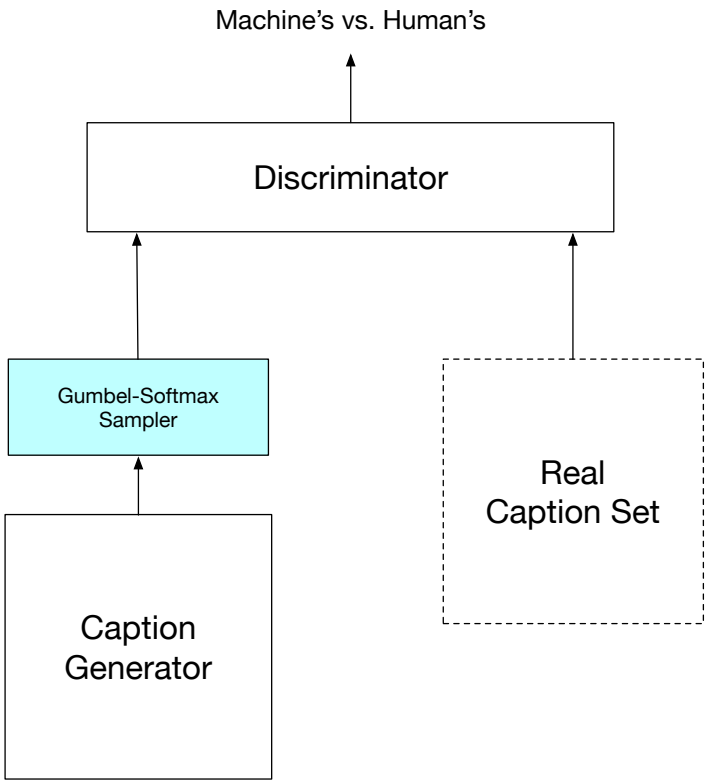


Model Architecture

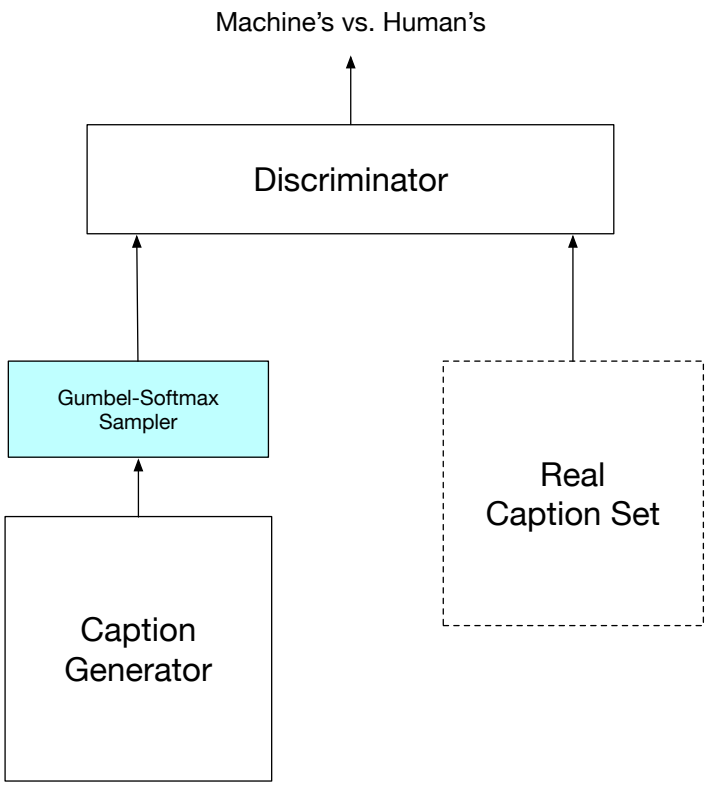


Model Architecture

Model Architecture



Model Architecture



Computational Data Flow

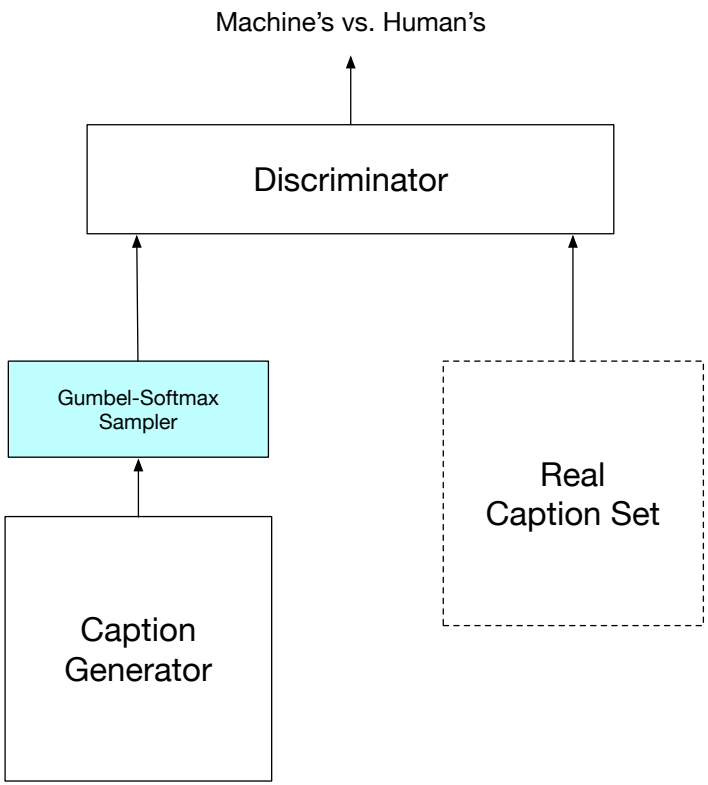
$$\text{hard} = \text{one\_hot}\left[\arg\max_i (g_i + \log\theta_i)\right]$$

$$\text{soft} = \text{softmax}\left(g + \log\theta\right)$$

$$\Pi_t = (\text{hard} - \text{soft}).\text{detach} + \text{soft}$$

$$w_t = W_e \Pi_t$$

Model Architecture



Computational Data Flow

$$\text{hard} = \text{one\_hot}\left[\arg\max_i(g_i + \log\theta_i)\right]$$

$$\text{soft} = \text{softmax}\left(g + \log\theta\right)$$

$$\Pi_t = (\text{hard} - \text{soft}).\text{detach} + \text{soft}$$

$$w_t = W_e\Pi_t$$

Equation Notations

$$\theta = \text{Categorical Distribution}$$

$$g_i \sim \text{Gumbel Distribution}$$

Loss Functions

Loss Functions

Teacher Forcing

$$L_{XE}(\theta) = -\sum_{t=1}^{T} \log(p_\theta(y_t^*|y_{1:t-1}^*))$$

Loss Functions

Teacher Forcing

$$L_{XE}(\theta) = -\sum_{t=1}^{T} \log(p_\theta(y_t^*|y_{1:t-1}^*))$$

CIDEr based REINFORCE (Rennie et al. 2016)

$$L_R(\theta) = -\mathbb{E}_{y_{1:T}^s \sim p_\theta}[r(y_{1:T}^s)]$$

$$\nabla_\theta(\theta) \approx -(r(y_{1:T}^s) - r(\hat{y}_{1:T}))\nabla_\theta \log p_\theta(y_{1:T}^s)$$

Loss Functions

Teacher Forcing

$$L_{XE}(\theta) = -\sum_{t=1}^{T} \log(p_\theta(y_t^* | y_{1:t-1}^*))$$

CIDEr based REINFORCE (Rennie et al. 2016)

$$L_R(\theta) = -\mathbb{E}_{y_{1:T}^s \sim p_\theta}[r(y_{1:T}^s)]$$

$$\nabla_\theta(\theta) \approx -(r(y_{1:T}^s) - r(\hat{y}_{1:T}))\nabla_\theta \log p_\theta(y_{1:T}^s)$$

Adversarial Loss

$$L_G(\theta) = -\mathbb{E}_{y_{1:T} \sim p_\theta}[D(y_{1:T})]$$

Experiments and Results

Experiments and Results

Table 1: A summary of the evaluation metrics used for image captioning

Metric	**Proposed to evaluate**	**Intuition**
BLEU	Machine translation	n-gram precision
ROUGE	Document summarization	n-gram recall
METEOR	Machine translation	n-gram with synonym matching
CIDEr	Image description generation	tf-idf weighted n-gram similarity

Experiments and Results

Table 1: A summary of the evaluation metrics used for image captioning

Metric	**Proposed to evaluate**	**Intuition**
BLEU	Machine translation	n-gram precision
ROUGE	Document summarization	n-gram recall
METEOR	Machine translation	n-gram with synonym matching
CIDEr	Image description generation	tf-idf weighted n-gram similarity

Table 2: Single-model performance on the MSCOCO Karpathy test

Training Method	**Model**	**BLEU-1**	**BLEU-4**	**METEOR**	**ROUGE-L**	**CIDEr**
REINFORCE CIDEr	SCST	-	0.333	0.263	**0.553**	1.114
	Bottom-up	0.766	0.340	0.265	0.549	1.111
	Ours	**0.767**	**0.342**	**0.266**	0.550	**1.117**
GANs	Ours	**0.770**	**0.345**	**0.269**	**0.554**	**1.121**

Experiments and Results

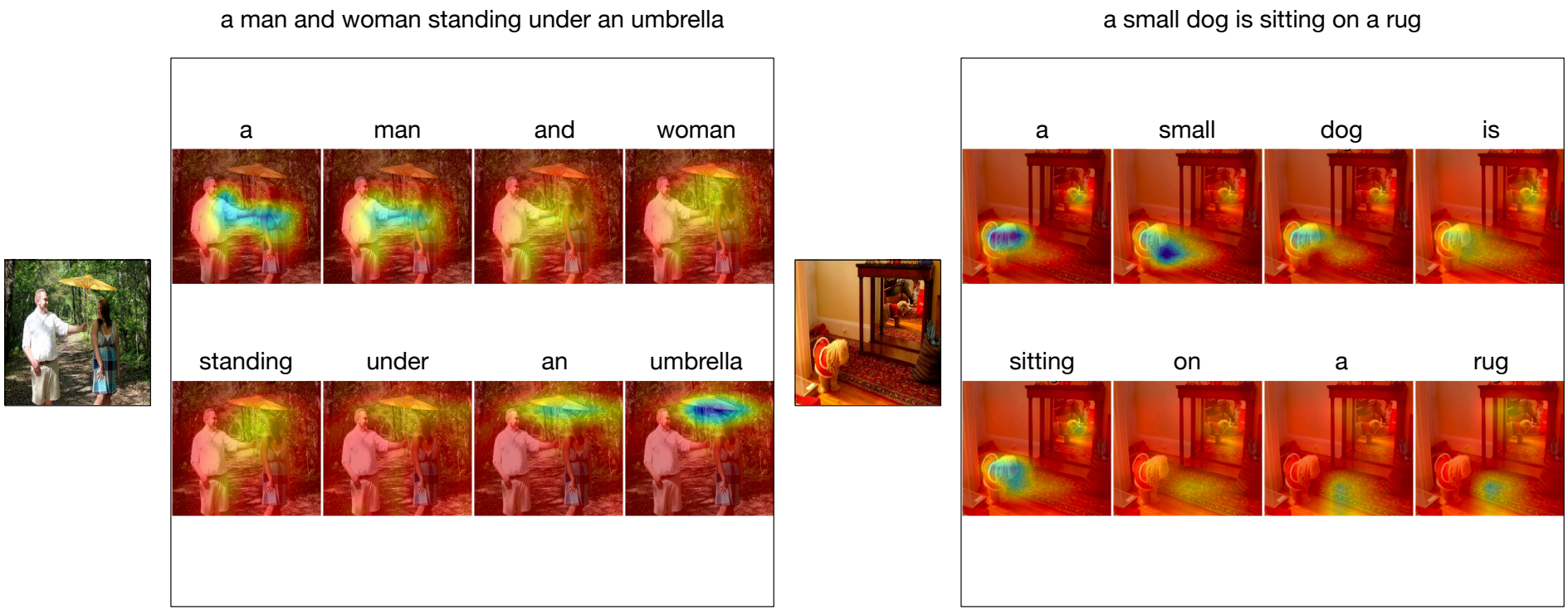


Figure 1: Qualitative captioning examples with first attention layer visualized.

Experiments and Results

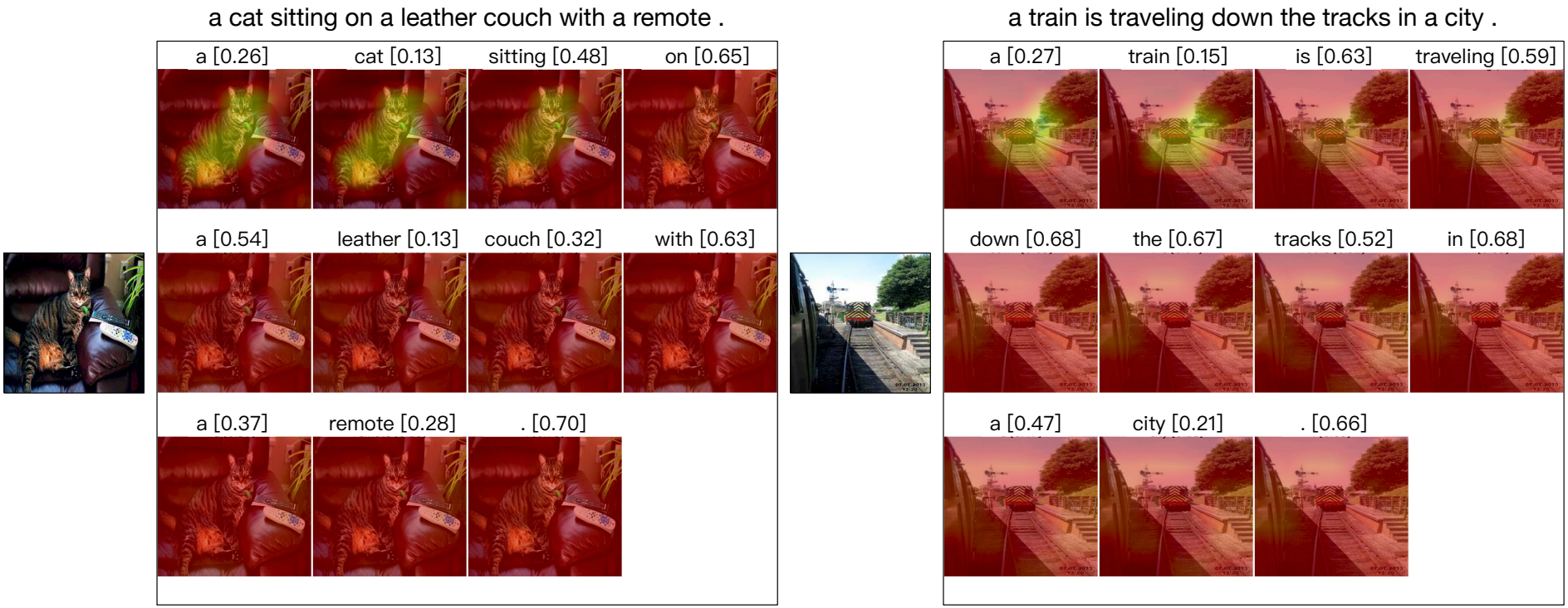


Figure 2: Captioning examples with sentinel (in brackets) and visual attentions.

Experiments and Results

Our Mixed Model: a man on a skateboard riding down the street

Our GANs Model: a man on a skateboard is riding down the ramp .

Our Mixed Model: a baseball player is swinging a bat at a ball .

Our GANs Model: a baseball player in orange shirt is kneeling on the field .

Figure 3: Caption examples of our mixed captioning model vs. our GANs model.

Overview

1. Introduction
   - Motivation
   - Problem Statement
   - Deep Learning Backgrounds
   - Hypotheses
   - Research Questions
2. Improved Image Captioning with Adversarial Loss
   - Model Architecture
   - Optimizations and Experiment Results
3. Congruence Measure between Image and Sentences
   - Pseudo Supervised Training
   - Loss Functions and Preliminary Results
4. Image Aspect Mining
   - Task Descriptions and Approaches
5. Conclusions and Research Timeline
6. Acknowledgements

Matching Sentences to Image

Matching Sentences to Image

Research Question

How can we measure the congruence of sentence-image pair if there is no supervised label for model training?

Matching Sentences to Image

Research Question

How can we measure the congruence of sentence-image pair if there is no supervised label for model training?

Relevance of a sentence to an image?

Matching Sentences to Image

Research Question

How can we measure the congruence of sentence-image pair if there is no supervised label for model training?

Relevance of a sentence to an image?

Review Text:

I tried this place for the first time today. I had the Sadie. It was great. Loved the pulled pork and beans. Just wish I had more BBQ sauce. I will go back again.

Ranked Sentences:

1. Loved the pulled pork and beans.
2. I had the Sadie.
3. It was great.
4. Just wish I had more BBQ sauce.
5. I tried this place for the first time today.
6. I will go back again.

Matching Sentences to Image

Matching Sentences to Image

Impacts

- Differentiate between text content and image description.
- Direct application for better image retrieval.
- Provide supervised label for image captioning.

Matching Sentences to Image

Impacts

- Differentiate between text content and image description.
- Direct application for better image retrieval.
- Provide supervised label for image captioning.

Related Works

- Text-to-image co-reference (Kong et al. 2014)
- Text query based video retrieval (Lin et al. 2014)
- Attribute grounding (Matuszek et al. 2012)

Matching Sentences to Image

Impacts

- Differentiate between text content and image description.
- Direct application for better image retrieval.
- Provide supervised label for image captioning.

Related Works

- Text-to-image co-reference (Kong et al. 2014)
- Text query based video retrieval (Lin et al. 2014)
- Attribute grounding (Matuszek et al. 2012)

Challenges

- No labeled review dataset for supervised learning.
- How to evaluate trained model quantitatively?

Pseudo Supervised Training

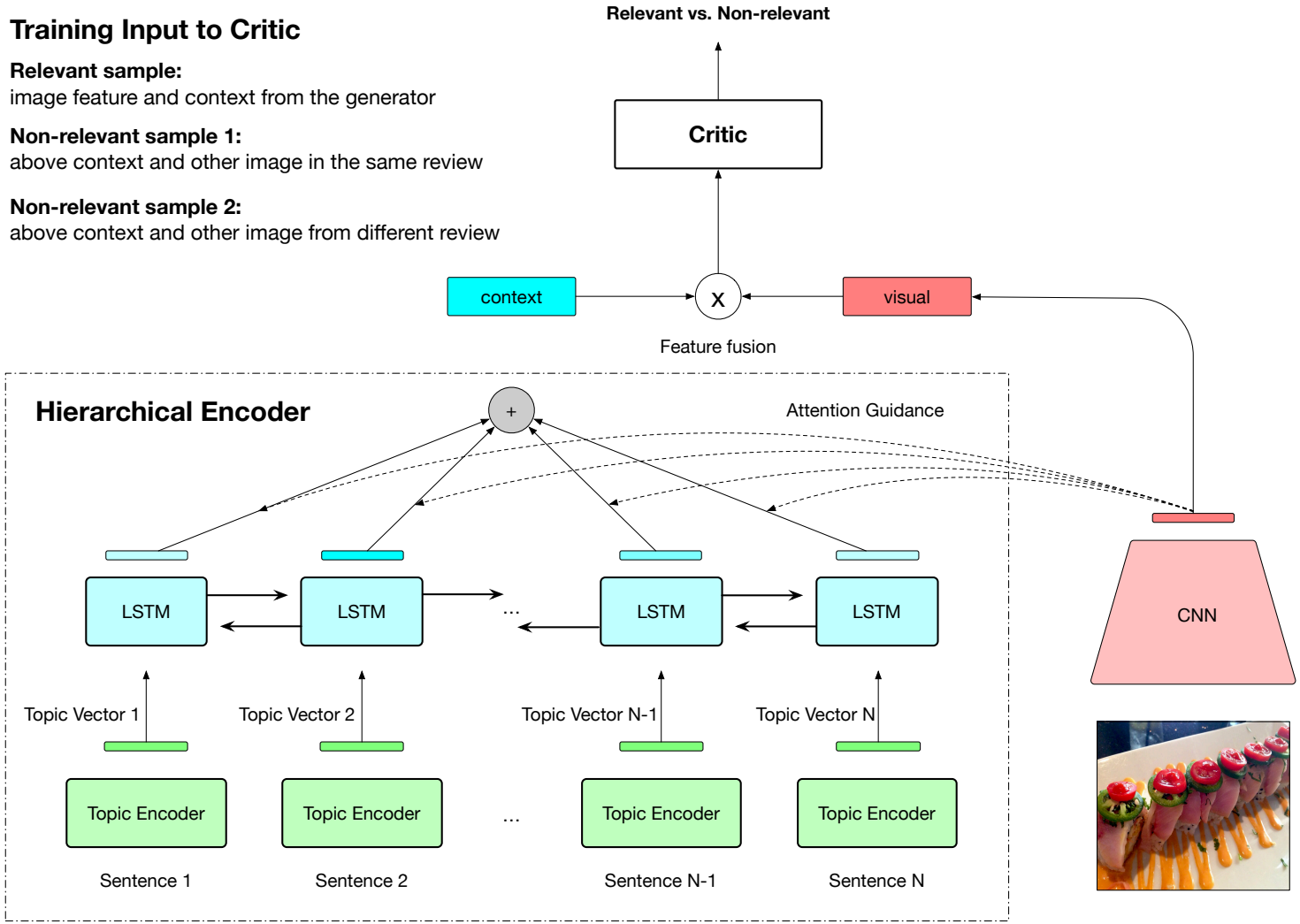
Pseudo Supervised Training

Training Input to Critic

Relevant sample:
image feature and context from the generator

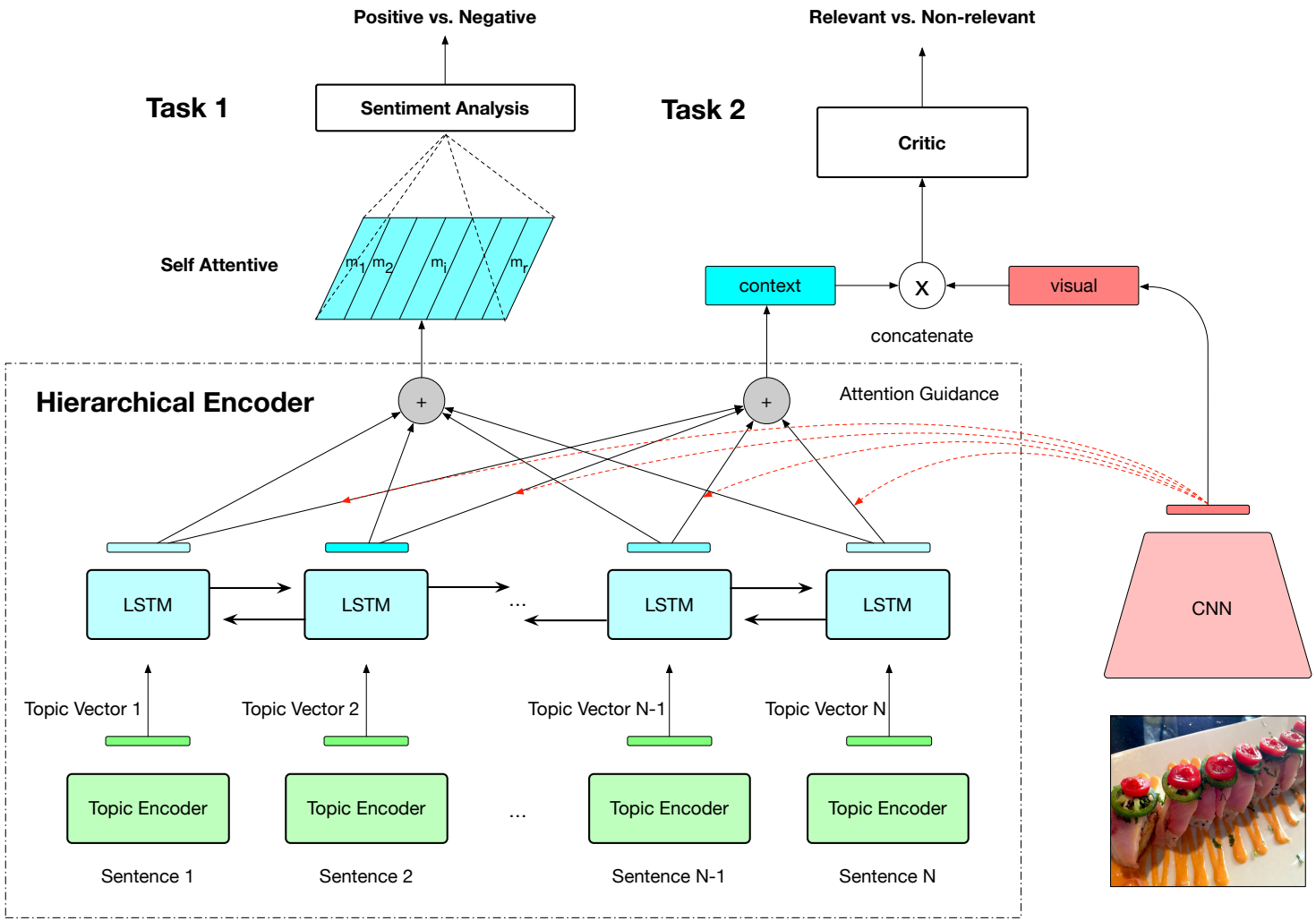
Non-relevant sample 1:
above context and other image in the same review

Non-relevant sample 2:
above context and other image from different review



Multi-task Learning

Multi-task Learning



Loss Function

Loss Function

New Relevance Loss

$$L(\theta; T, I) = \sum_{k=1}^{M} [- f(\mathrm{CNN}(I_{t=k}), g(T, I_{t=k})) + f(\mathrm{CNN}(I_{t\neq k}), g(T, I_{t=k}))]$$

Loss Function

New Relevance Loss

$$L(\theta; T, I) = \sum_{k=1}^{M} [- f(\mathrm{CNN}(I_{t=k}), g(T, I_{t=k})) + f(\mathrm{CNN}(I_{t\neq k}), g(T, I_{t=k}))]$$

Self Attention Loss

$$A = \mathrm{softmax}(W_{s2}\tanh(W_{s1}H^{\top} + \cancel{\text{outsidesignal}}))$$

$$L_{\mathsf{sent}} = L_{XE} + \lambda_1 ||A^{\top}A - I||_F$$

Loss Function

New Relevance Loss

$$L(\theta; T, I) = \sum_{k=1}^{M} [- f(\text{CNN}(I_{t=k}), g(T, I_{t=k})) + f(\text{CNN}(I_{t\neq k}), g(T, I_{t=k}))]$$

Self Attention Loss

$$A = \text{softmax}(W_{s2}\tanh(W_{s1}H^{\top} + \cancel{\text{outsidesignal}}))$$

$$L_{\text{sent}} = L_{XE} + \lambda_1 ||A^{\top}A - I||_F$$

Multi-task Learning Loss

$$L = L(\theta; T, I) + \lambda_2 L_{\text{sent}}$$

Preliminary Results

Preliminary Results

Dataset: Yelp restaurant reviews with $13,500$ samples.

	Review Text	Top 3 Sentences with Basic Model	Top 3 Sentences with Multi-task Model
	Want a great, no frills taco? This is your place. The salsa bar is phenomenal too! I had the el Capitan with asada. It was delicious, but a little salty. My boyfriend had the barbacoa, tripe, and lengua tacos, and decided the tripe was his favorite. Great location and great place!	1. My boyfriend had the barbacoa, tripe, and lengua tacos, and decided the tripe was his favorite. (0.132) 2. Great location and great place! (0.121) 3. It was delicious, but a little salty. (0.117)	1. Want a great, no frills taco? (0.144) 2. This is your place.(0.143) 3. The salsa bar is phenomenal too! (0.143)
	I tried this place for the first time today. I had the Sadie. It was great. Loved the pulled pork and beans. Just wish I had more BBQ sauce. I will go back again.	1. Loved the pulled pork and beans. (0.189) 2. I will go back again. (0.168) 3. It was great. (0.166)	1. Loved the pulled pork and beans. (0.191) 2. I had the Sadie. (0.171) 3. It was great. (0.167)

Future Work and Evaluation Plan

Future Work and Evaluation Plan

Dataset Labeling

- Yelp restaurant dataset from 17 U.S. major cities.
- $1,000$ reviews manually tagged, more being done.

Future Work and Evaluation Plan

Dataset Labeling

- Yelp restaurant dataset from 17 U.S. major cities.
- $1,000$ reviews manually tagged, more being done.

Quantitative Evaluation Metric

- Average Precision@n, Recall@n, ...

Future Work and Evaluation Plan

Dataset Labeling

- Yelp restaurant dataset from 17 U.S. major cities.
- $1,000$ reviews manually tagged, more being done.

Quantitative Evaluation Metric

- Average Precision@n, Recall@n, ...

Baseline

- Relate image captions to review sentences.

Overview

1. Introduction
   - Motivation
   - Problem Statement
   - Deep Learning Backgrounds
   - Hypotheses
   - Research Questions
2. Improved Image Captioning with Adversarial Loss
   - Model Architecture
   - Optimizations and Experiment Results
3. Congruence Measure between Image and Sentences
   - Pseudo Supervised Training
   - Loss Functions and Preliminary Results
4. Image Aspect Mining
   - Task Descriptions and Approaches
5. Conclusions and Research Timeline
6. Acknowledgements

Image Aspect Mining

Image Aspect Mining

What is Image Aspect Mining?

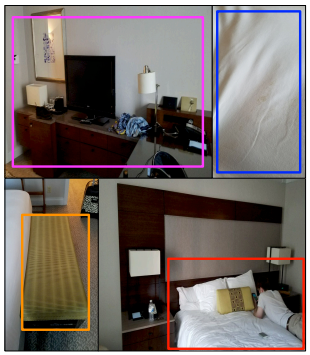
What is Image Aspect Mining?

So so. after a week the refrigerator is working fine. the box was dented and had holes in it, and the refrigerator had dents and scratched that matched the damage to the box.

box: dented, holes

refrigerator: dents, scratched

Amazon Review



I was utterly disappointed with my experience here. Literally nothing about this hotel was luxurious in the slightest. Our room was completely lackluster, outdated and even dirty. The sheets were stained and the ancient foot stool at the end of the bed looked filthy. Service was fine, bed was comfortable, and bathroom was clean. Those are the only positives I can list.

Room: lackluster, outdated, dirty

Bed: comfortable

Sheets: ancient, stained

Foot stool: filthy

Yelp Hotel Review

Image Aspect Mining

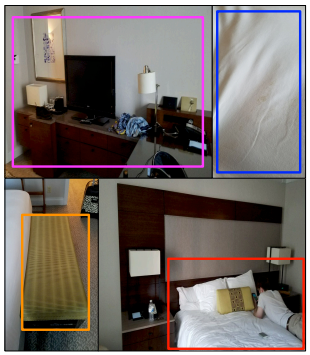
What is Image Aspect Mining?

So so. after a week the refrigerator is working fine. the box was dented and had holes in it, and the refrigerator had dents and scratched that matched the damage to the box.

box: dented, holes

refrigerator: dents, scratched

Amazon Review



I was utterly disappointed with my experience here. Literally nothing about this hotel was luxurious in the slightest. Our room was completely lackluster, outdated and even dirty. The sheets were stained and the ancient foot stool at the end of the bed looked filthy. Service was fine, bed was comfortable, and bathroom was clean. Those are the only positives I can list.

Room: lackluster, outdated, dirty

Bed: comfortable

Sheets: ancient, stained

Foot stool: filthy

Yelp Hotel Review

Image Aspect Mining

- Recognize topical aspects in review text.
- Attend to specific regions in images.
- Infer fine-grained ratings.

Image Aspect Mining

Image Aspect Mining

Impacts

- Aspect ratings for merchants to provide better service, and consumers to target specific requirements.
- Identify aspect rating congruence between text and images.

Image Aspect Mining

Impacts

- Aspect ratings for merchants to provide better service, and consumers to target specific requirements.
- Identify aspect rating congruence between text and images.

Related Works

- Object detection: R-CNN families, YOLO (Girshick et al. 2014, Redmon et al. 2016)
- Image-text feature fusion: MUTAN (Ben et al. 2017)
- Multi-step attention: stacked attention (Yang et al. 2016)

Approaches

Approaches

Baseline

- Image Aspect: proposals from R-CNN/YOLO;
- Text: RNN - LSTM;
- Feature fusion: concatenation/element-wise product;

Approaches

Baseline

- Image Aspect: proposals from R-CNN/YOLO;
- Text: RNN - LSTM;
- Feature fusion: concatenation/element-wise product;

Bilinear + Attention

- Bilinear feature fusion: $y = x_1 * A * x_2$;
- Stacked attentions: multi-step glimpses;

Approaches

Baseline

- Image Aspect: proposals from R-CNN/YOLO;
- Text: RNN - LSTM;
- Feature fusion: concatenation/element-wise product;

Bilinear + Attention

- Bilinear feature fusion: $y = x_1 * A * x_2$;
- Stacked attentions: multi-step glimpses;

Generative

- Text2image: Generative Adversarial Networks;
- Gumbel-Softmax, Adversarially Regularized Autoencoder (Kim et al. 2017);

Evaluations

Evaluations

Metric Definition

$$\Delta^2_{\text{aspect}} = \sum_{d=1}^{|D|} \sum_{i=1}^{k} (s_{di} - s^*_{di})^2 / (k \times |D|).$$

Evaluations

Metric Definition

$$\Delta^2_{\text{aspect}} = \sum_{d=1}^{|D|} \sum_{i=1}^{k} (s_{di} - s^*_{di})^2 / (k \times |D|).$$

- Review corpus D and k pre-defined aspects;
- s_{di}: predicted rating on aspect i in review d;
- s^*_{di}: ground truth rating on aspect i in review d;

Evaluations

Metric Definition

$$\Delta^2_{\text{aspect}} = \sum_{d=1}^{|D|} \sum_{i=1}^{k} (s_{di} - s^*_{di})^2 / (k \times |D|).$$

- Review corpus D and k pre-defined aspects;
- s_{di}: predicted rating on aspect i in review d;
- s^*_{di}: ground truth rating on aspect i in review d;

Datasets

- TripAdvisor hotel reviews: aspect level ratings provided
- Yelp restaurant reviews
- Amazon reviews

Overview

1. Introduction
   - Motivation
   - Problem Statement
   - Deep Learning Backgrounds
   - Hypotheses
   - Research Questions
2. Improved Image Captioning with Adversarial Loss
   - Model Architecture
   - Optimizations and Experiment Results
3. Congruence Measure between Image and Sentences
   - Pseudo Supervised Training
   - Loss Functions and Preliminary Results
4. Image Aspect Mining
   - Task Descriptions and Approaches
5. Conclusions and Research Timeline
6. Acknowledgements

Conclusions

Conclusions

Image Captioning

- Neural image captioning models fine-tuned with GANs
- Improved quality and evaluation metric performance

Conclusions

Image Captioning

- Neural image captioning models fine-tuned with GANs
- Improved quality and evaluation metric performance

Matching Sentences to Image

- Pseudo supervised training for matching sentences to image
- Preliminary qualitative effectiveness

Conclusions

Image Captioning

- Neural image captioning models fine-tuned with GANs
- Improved quality and evaluation metric performance

Matching Sentences to Image

- Pseudo supervised training for matching sentences to image
- Preliminary qualitative effectiveness

Image Aspect Mining

- Proposed a new AI task
- Datasets, baseline models, and evaluation protocols

Research Timeline

Task	2017		2018		
	Summer	Fall	Spring	Summer	Fall
Topic 1 - Image Captioning with GANs	✓	✓	✓		
Basic image caption model setup	✓				
Training with GANs framework		✓			
Publish the results			✓		
Topic 2 - Matching Sentences to Images			✓	✓	
Model design and experiments			✓		
Dataset labeling and evaluation			✓		
Publish the results				✓	
Topic 3 - Image Aspect Mining				✓	✓
Dataset collection and labeling				✓	
Model implementation and evaluation				✓	
Publish the results					✓
Writing the dissertation					✓
Research defense					✓
Final defense					✓

Publications

[1]. **Yufeng Ma**, Zheng Xiang, Qianzhou Du, and Weiguo Fan, "Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning," *International Journal of Hospitality Management*, vol. 71, pp. 120–131, 2018.

[2]. **Yufeng Ma**, Tingting Jiang, Chandani Shrestha, Edward Alan Fox, Jian Wu, and C. Lee Giles, "Scenarios for advanced services in an ETD digital library," in *Proceedings of ETD2017, the 20th international symposium on electronic theses and dissertations, Washington, DC, August 7-9*, 2017.

[3]. Zheng Xiang, Qianzhou Du, **Yufeng Ma**, and Weiguo Fan, "Assessing reliability of social media data: Lessons from mining tripadvisor hotel reviews," in *Information and Communication Technologies in Tourism 2017*, pp. 625–638, Springer, 2017 **Best Research Paper Award**.

[4]. Zheng Xiang, Qianzhou Du, **Yufeng Ma**, and Weiguo Fan, "A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism," *Tourism Management*, vol. 58, pp. 51–65, 2017.

[5]. **Yufeng Ma**, Long Xia, Wenqi Shen, Mi Zhou, and Weiguo Fan, "A surrogate-based generic classifier for Chinese TV series reviews," *Information Discovery and Delivery*, vol. 45, no. 2, pp. 66–74, 2017.

[6]. Long Xia, **Yufeng Ma**, and Weiguo Fan, "VTIR at the NTCIR-12 2016 lifelog semantic access task," *Proceedings of NTCIR-12, Tokyo, Japan*, 2015.

Acknowledgements

Acknowledgements

Doctoral committee

Weiguo (Patrick) Fan, Edward A. Fox,
G. Alan Wang, Bert Huang, Zhongju (John) Zhang

Acknowledgements

Doctoral committee

Weiguo (Patrick) Fan, Edward A. Fox,
G. Alan Wang, Bert Huang, Zhongju (John) Zhang

Organizations and Projects

Center for Business Intelligence and Analytics (CBIA), Digital Library Research Laboratory (DLRL), Deposition Summarization Project

Acknowledgements

Doctoral committee

Weiguo (Patrick) Fan, Edward A. Fox,
G. Alan Wang, Bert Huang, Zhongju (John) Zhang

Organizations and Projects

Center for Business Intelligence and Analytics (CBIA), Digital Library Research Laboratory (DLRL), Deposition Summarization Project

Lab colleagues, friends and others

Xuan Zhang, Liuqing Li, Prashant Chandrasekar, Ziqian Song, Qianzhou Du, Qinxiang An, Zheng Xiang, Long Xia, Wenqi Shen, Yu Wang, Siyu Mi, Xinyue Wang, Yumin Dai, Sunshin Lee, Mohamed Magdy, Steve Hughes, Brian Kopczynski, Dhruv Batra, ...

Overview

1. Introduction
   - Motivation
   - Problem Statement
   - Deep Learning Backgrounds
   - Hypotheses
   - Research Questions
2. Improved Image Captioning with Adversarial Loss
   - Model Architecture
   - Optimizations and Experiment Results
3. Congruence Measure between Image and Sentences
   - Pseudo Supervised Training
   - Loss Functions and Preliminary Results
4. Image Aspect Mining
   - Task Descriptions and Approaches
5. Conclusions and Research Timeline
6. Acknowledgements

Thank You!
Questions & Comments?