



# Person Search: A New Research Paradigm

Shuang Li

The Chinese University of Hong Kong



# Person Re-identification



# Person Re-identification



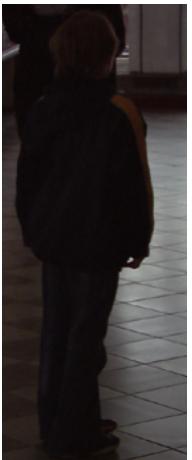
# Challenges



Pose variation

Viewpoints

Occlusion



Low illumination

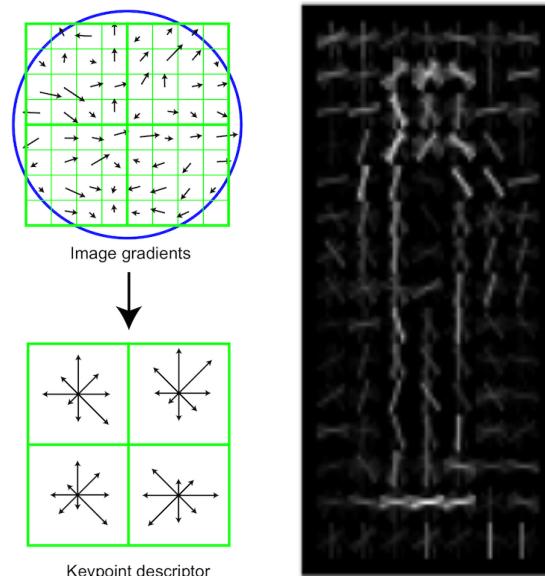
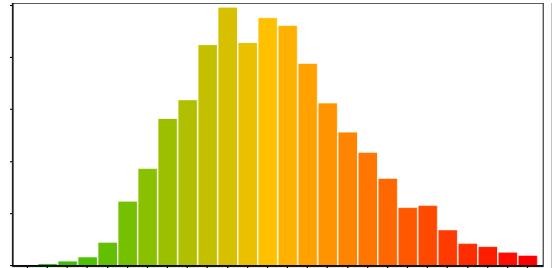
Low resolution

Background clutter

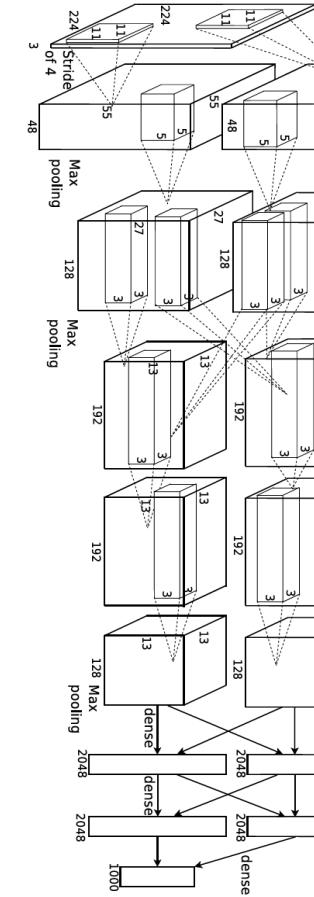
# Solutions

- Feature learning
- Metric learning

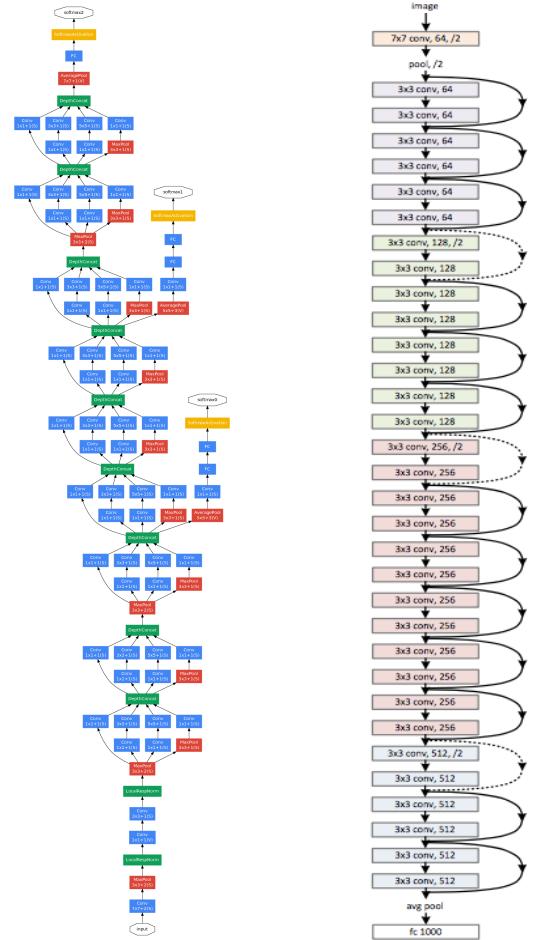
# Feature Learning



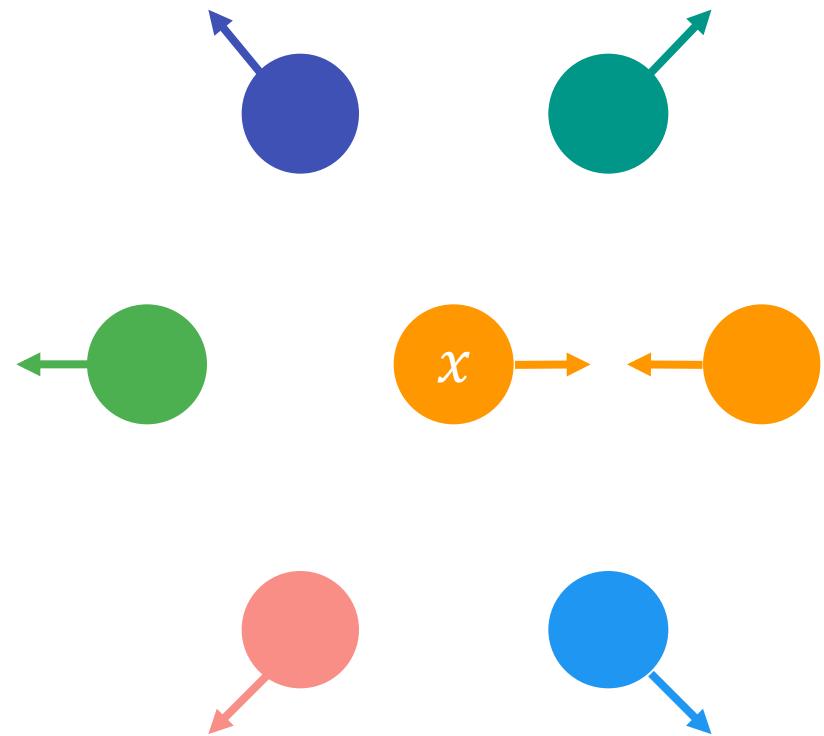
Hand-crafted features



Deep features



# Metric Learning



$$d(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j)$$

Minimize distance between  
same person

Maximize distances among  
different people

# Person Re-Identification



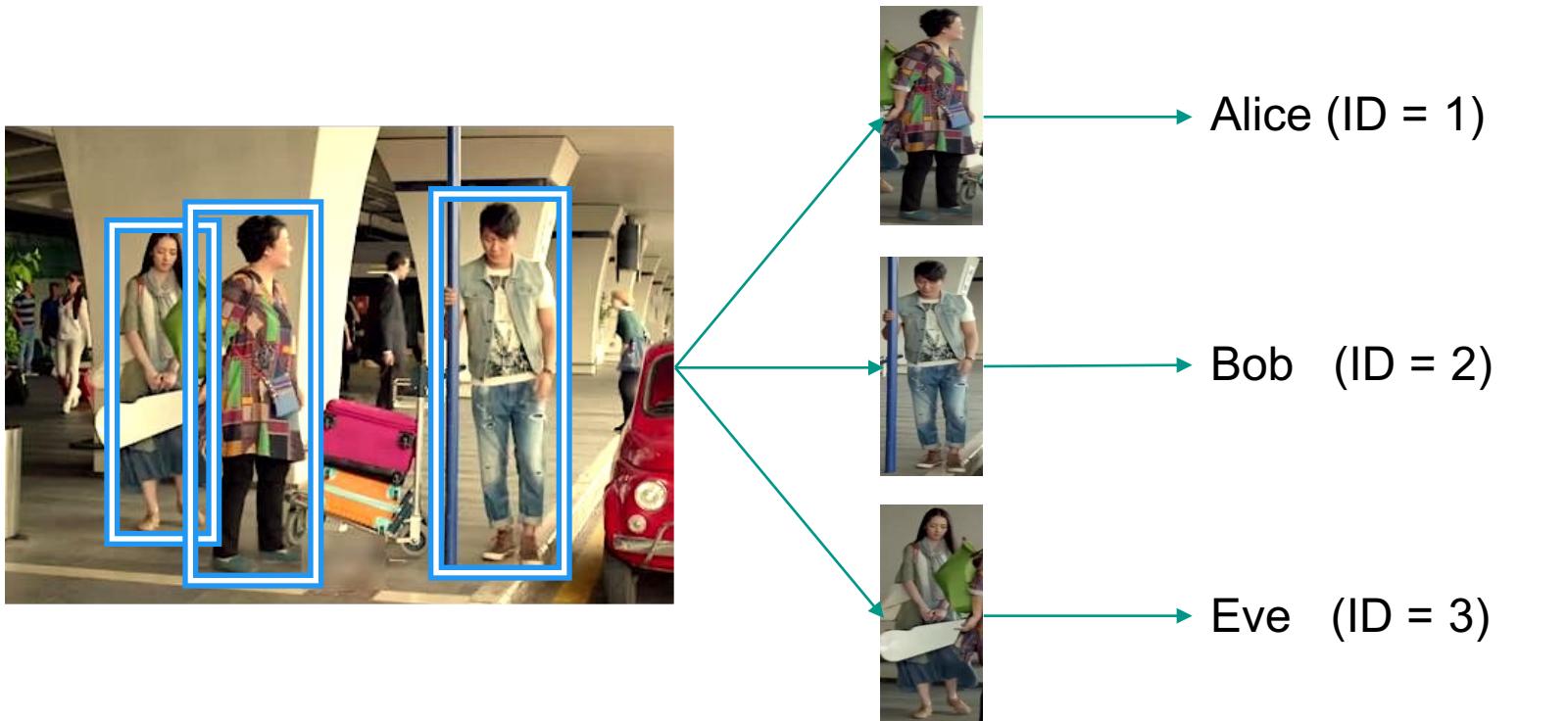
- Matching
- Manually cropped

# Person Search



- Finding
- Whole scene images

# Detection + Identification



Feature -  $x$

Category -  $y$

# Joint Detection and Identification Feature Learning for Person Search

CVPR  
2017

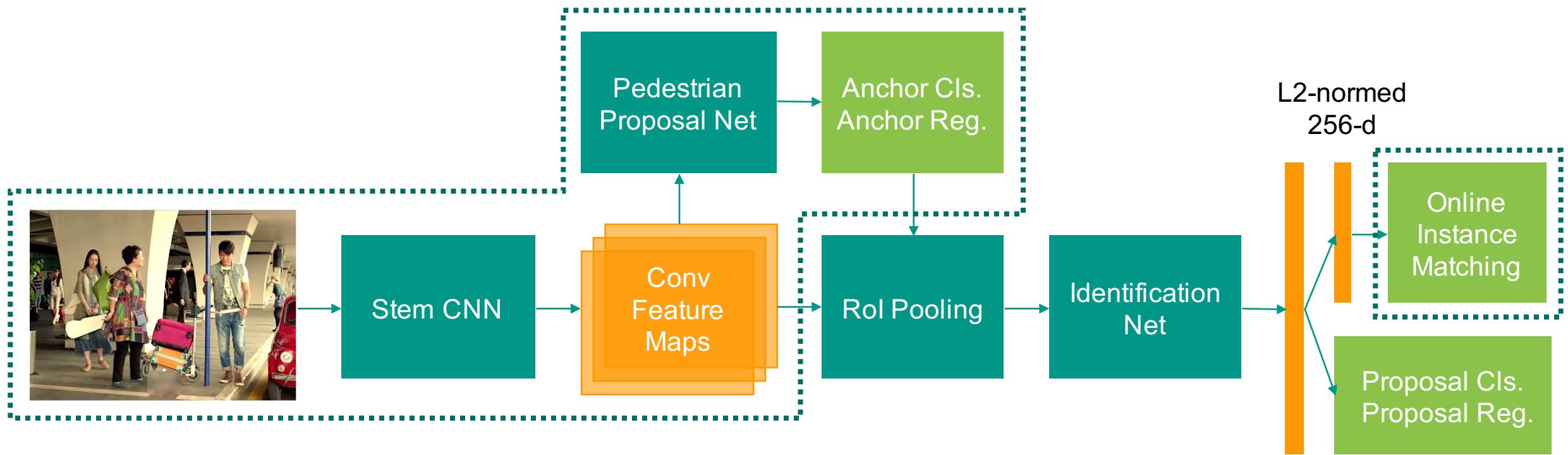
Tong Xiao\*, Shuang Li\*, Bochao Wang, Liang Lin, Xiaogang Wang



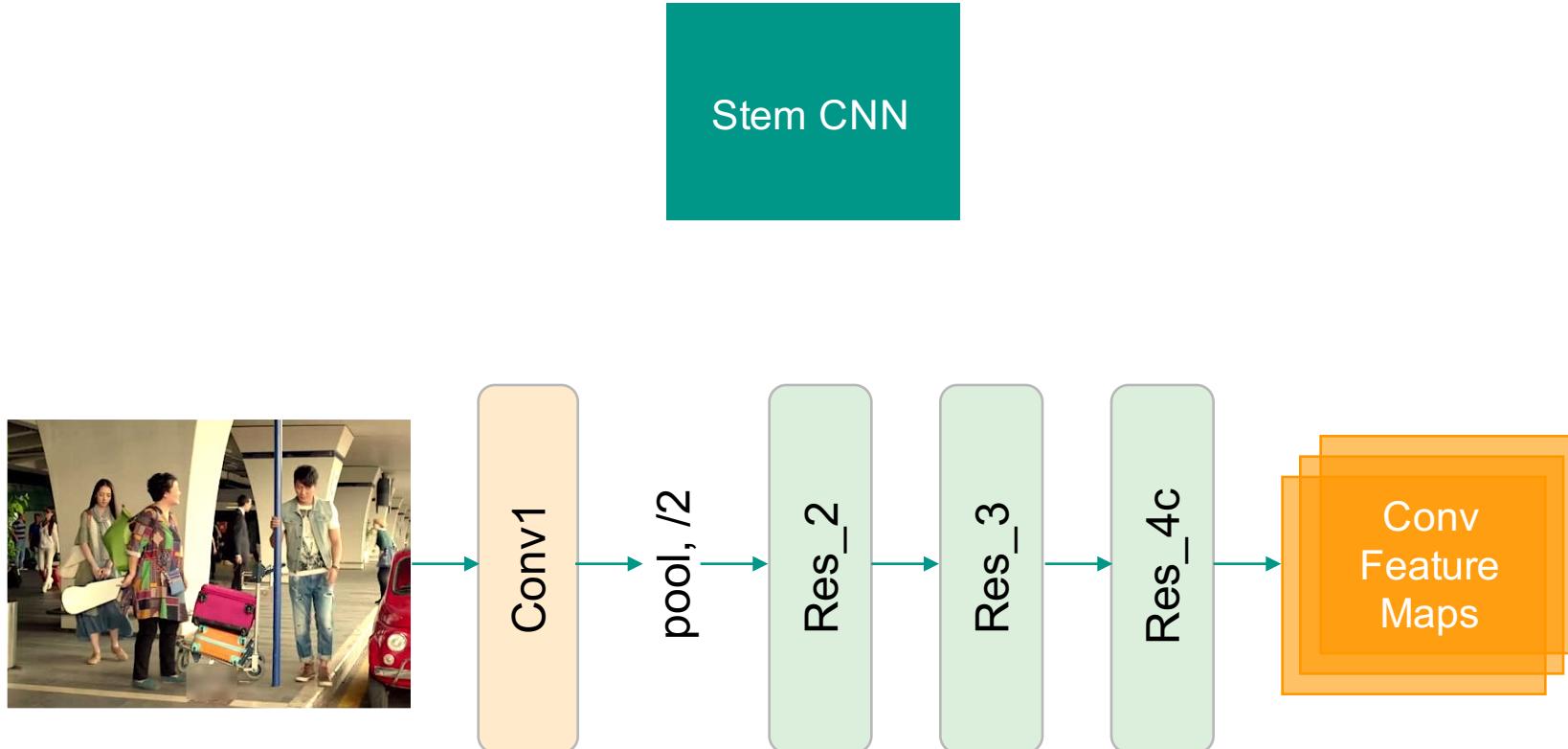
SENSETIM

\* Equal Contribution

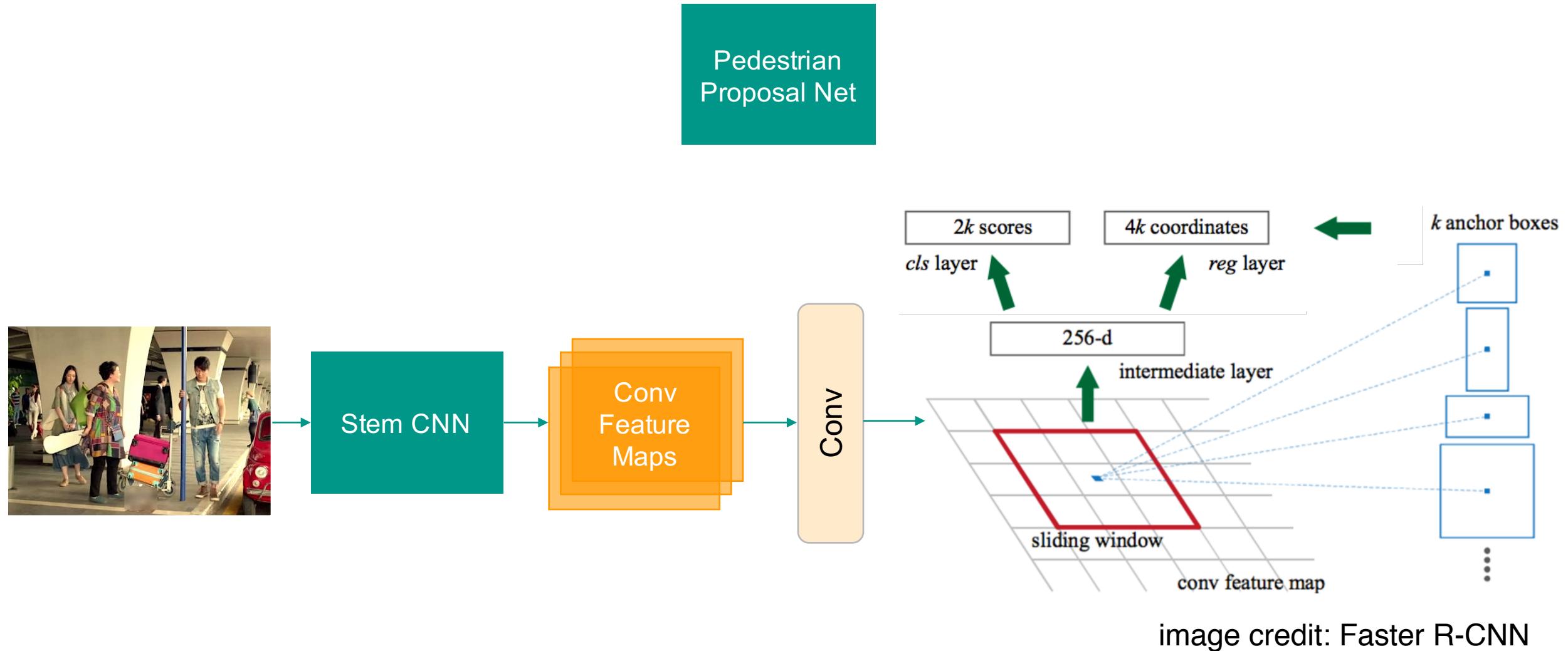
# Framework Overview



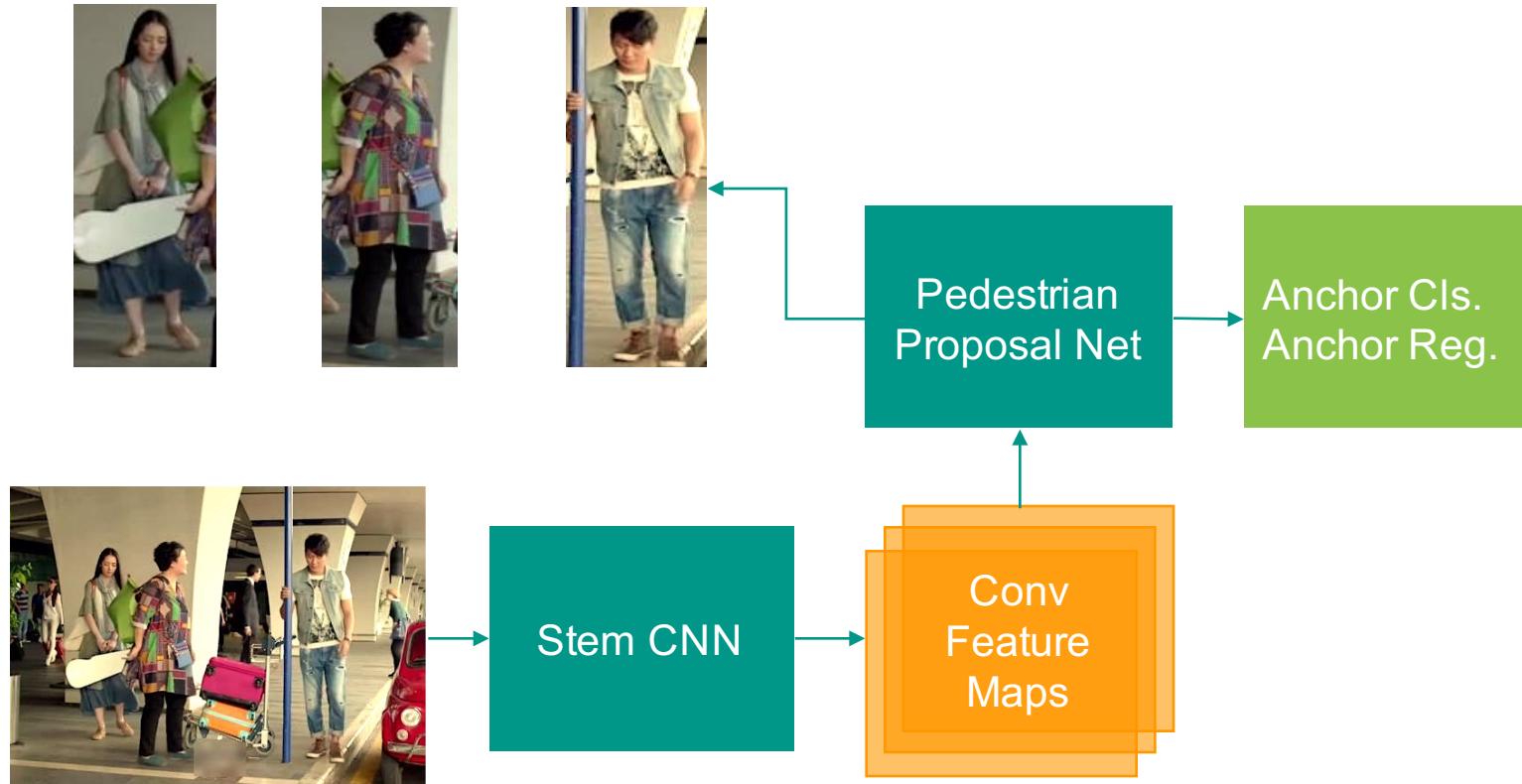
# Stem CNN



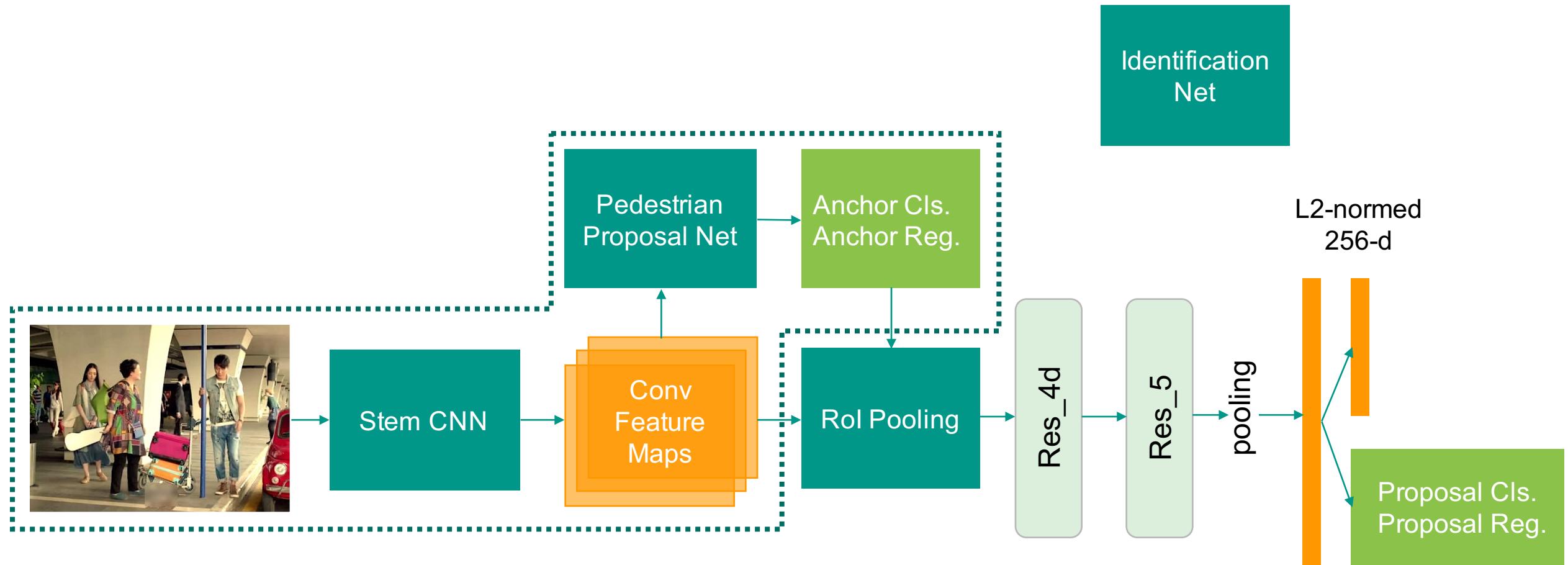
# Pedestrian Proposal Net



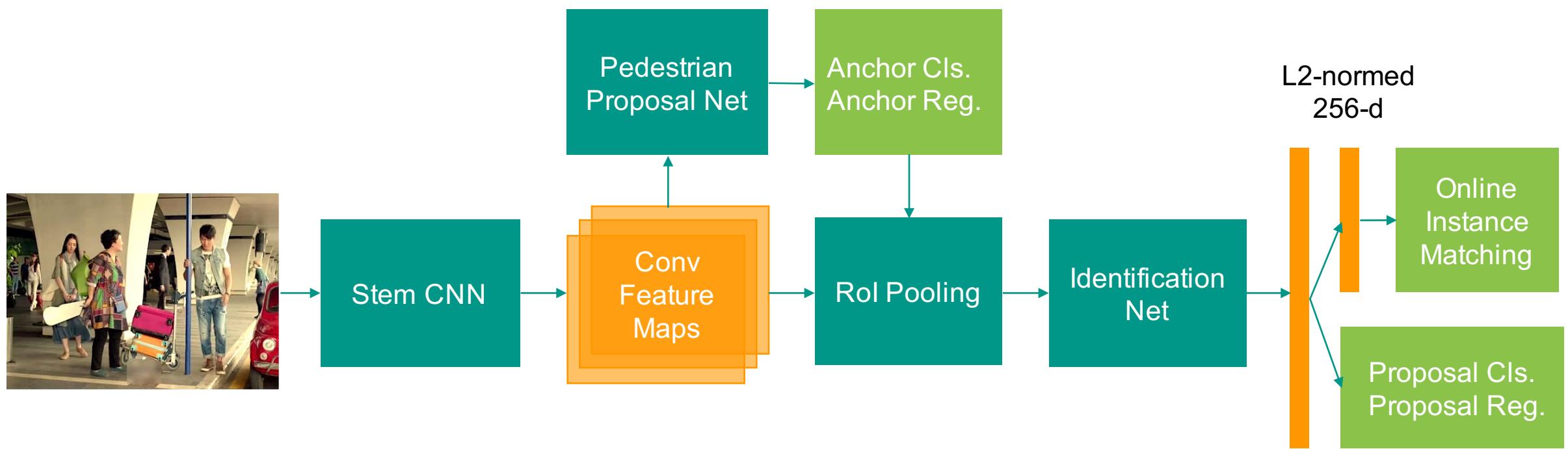
# Pedestrian Proposal Net



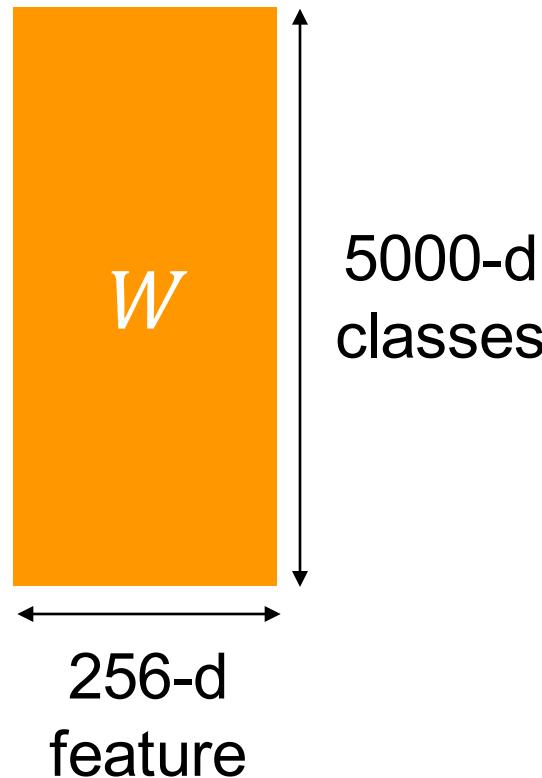
# Identification Net



# Joint Optimization

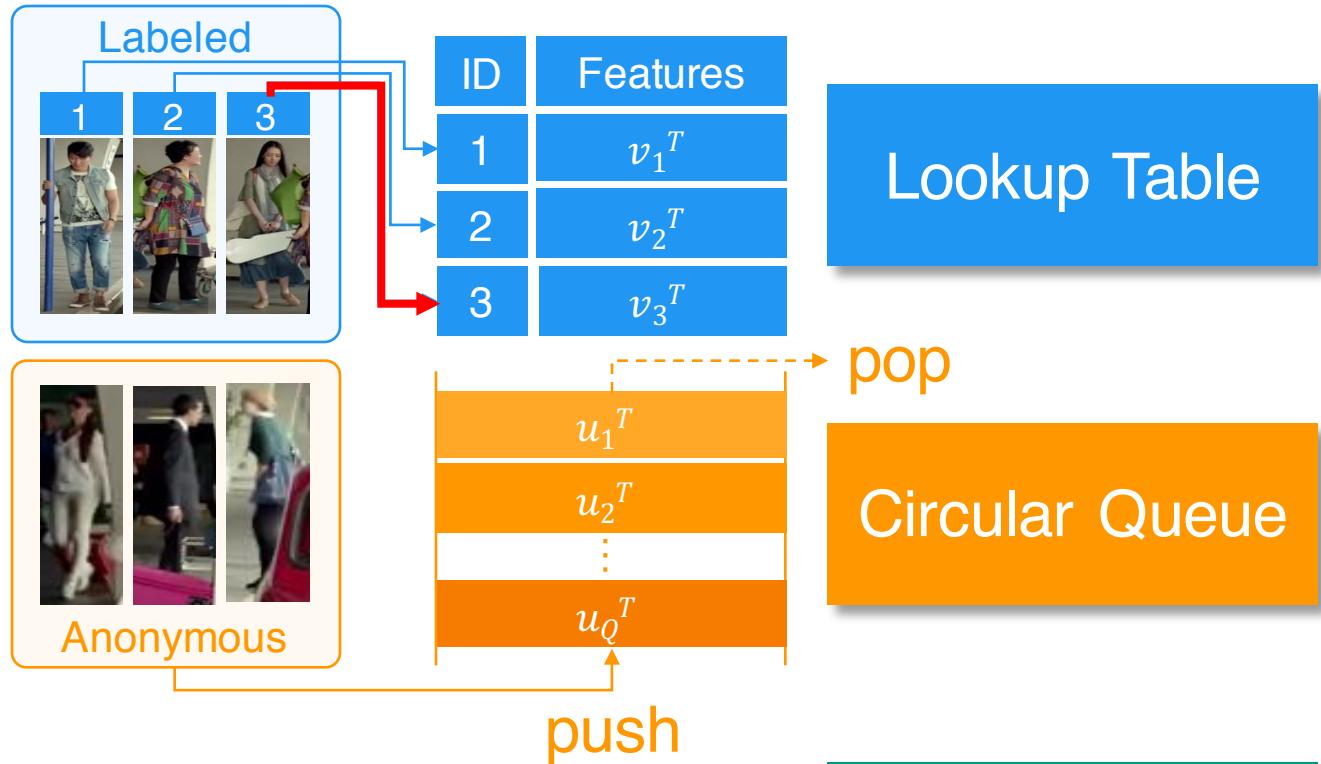


# Problems in Softmax Loss



- $p(y = i|x) = \frac{\exp(w_i^T x)}{\sum_j \exp(w_j^T x)}$
- # Positive classes  $\leq$  batch size  $\ll$  5000
- No positive samples for most rows
- $W$  cannot be learned effectively

# Online Instance Matching (OIM)



$$p(y = i|x) = \frac{\exp(v_i^T x)}{\sum_j \exp(v_j^T x) + \sum_k \exp(u_k^T x)}$$

No Parameters

# Experiments

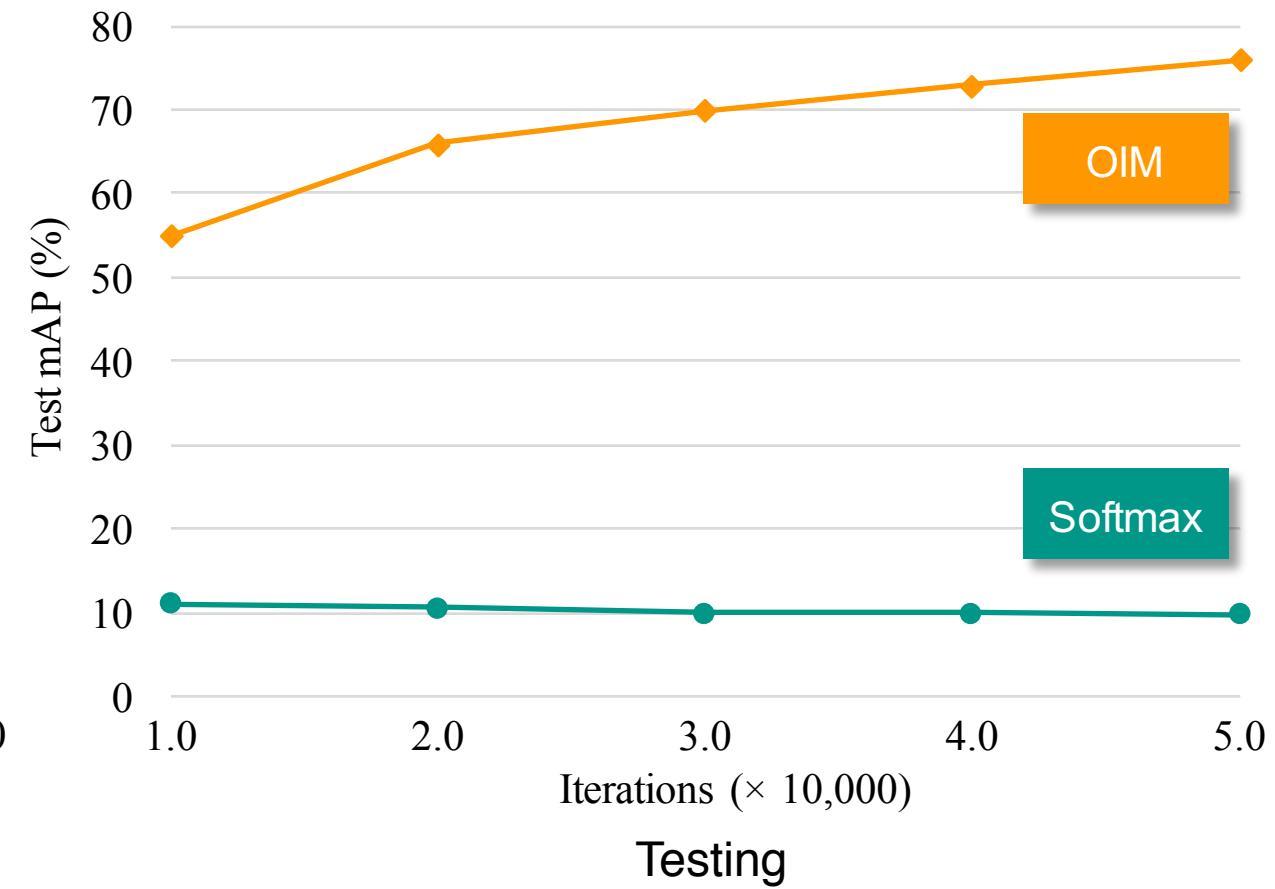
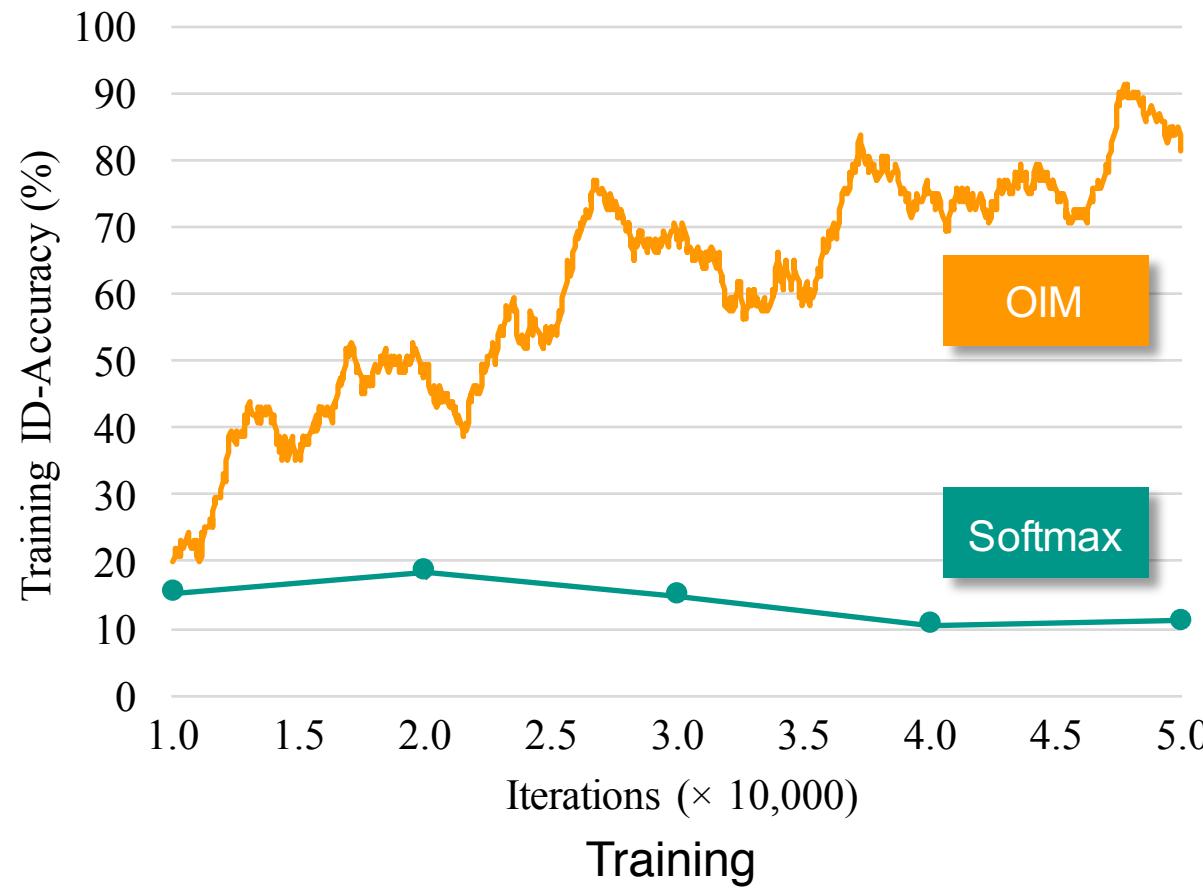
# OIM vs. Softmax Loss

8,432 Identities

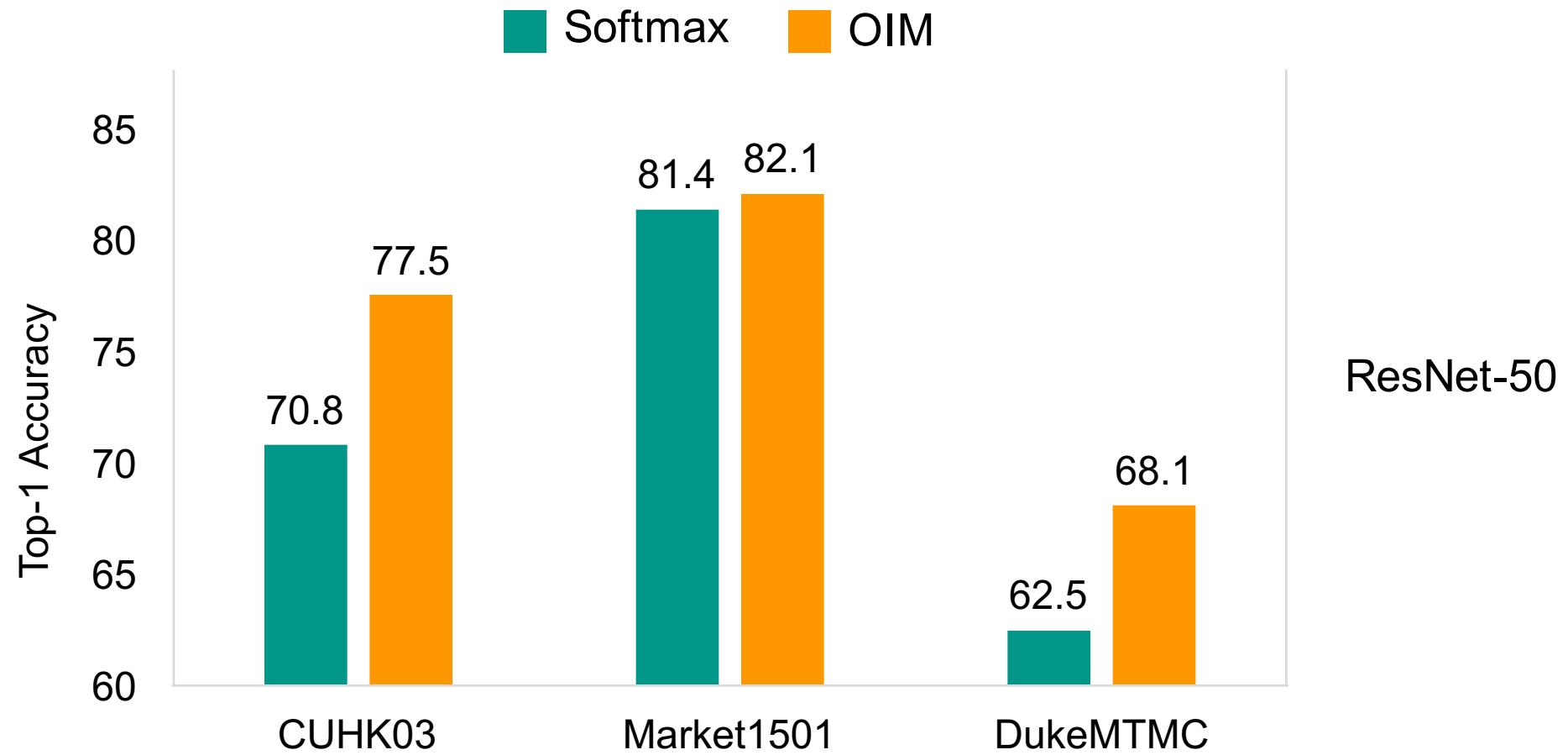
18,184 Images

96,143 Pedestrians

# OIM vs. Softmax Loss

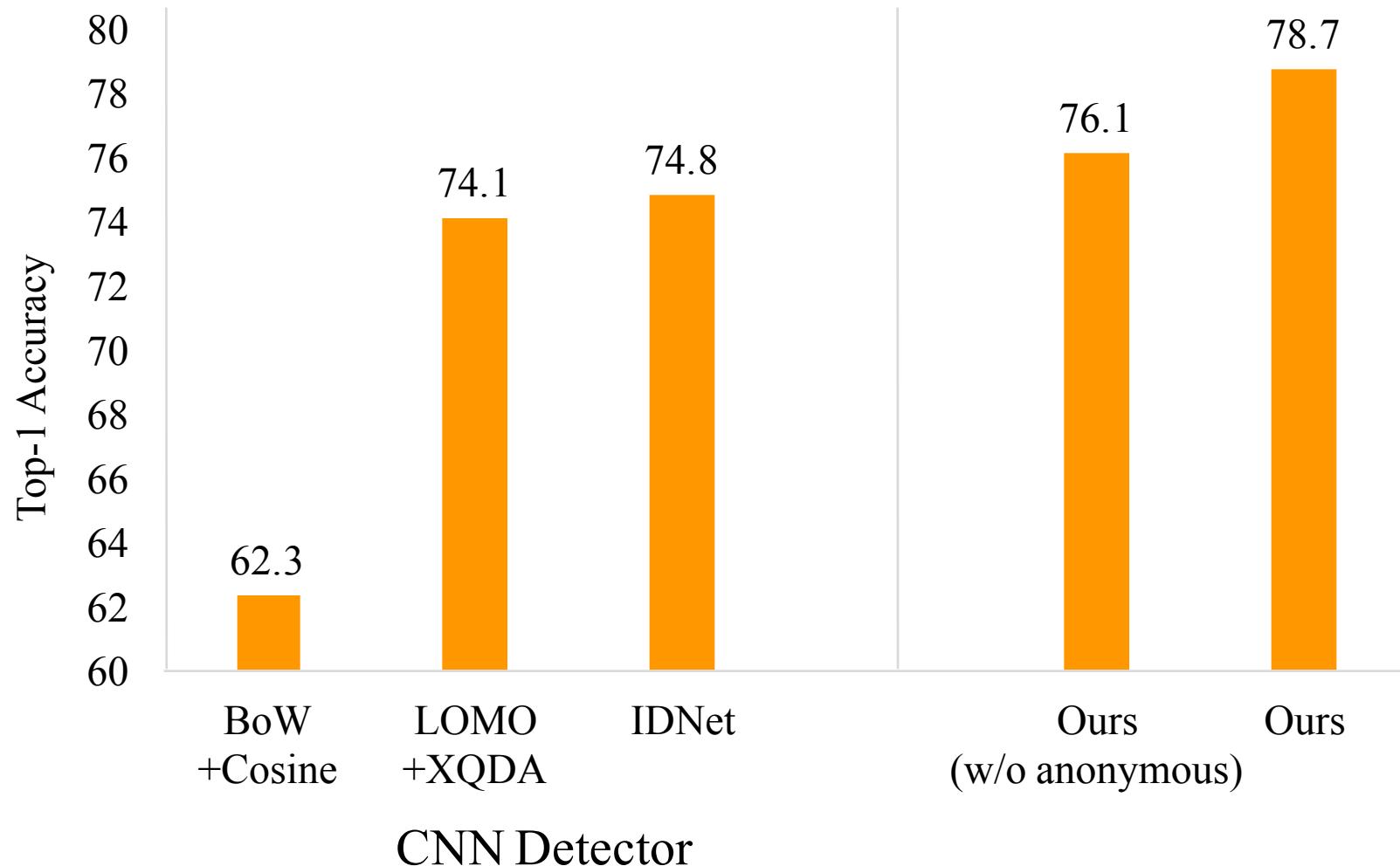


# OIM vs. Softmax Loss



Top-1 accuracy of using OIM or Softmax loss for standard person re-id task.

# Joint vs. Separate Detection + Identification



# Outcomes

- Propose a joint detection and identification framework
- Present an Online Instance Matching loss function
- Collect a large-scale dataset for person search

# Identity-Aware Textual-Visual Matching with Latent Co-attention

ICCV  
2017

Shuang Li, Tong Xiao, Hongsheng Li, Wei Yang, Xiaogang Wang

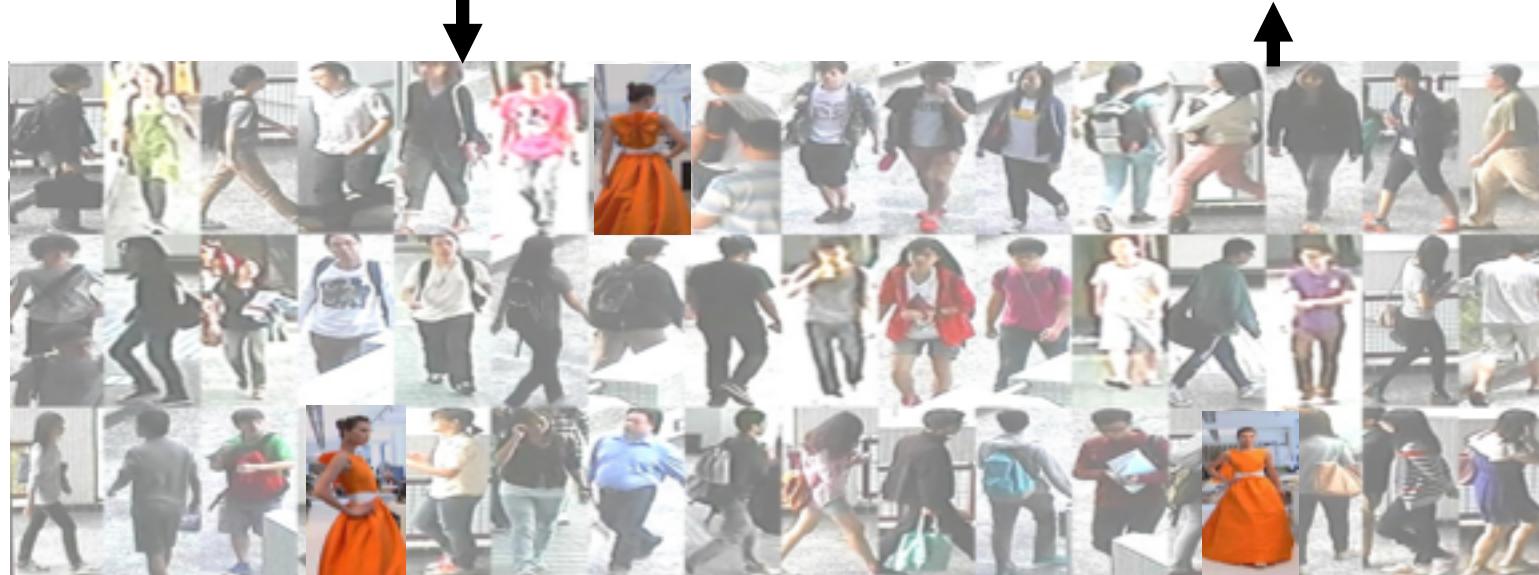


# Text based Person Search

## Query Description

The woman is wearing a long, bright orange gown with a white belt at her waist. She has her hair pulled back into a bun or ponytail.

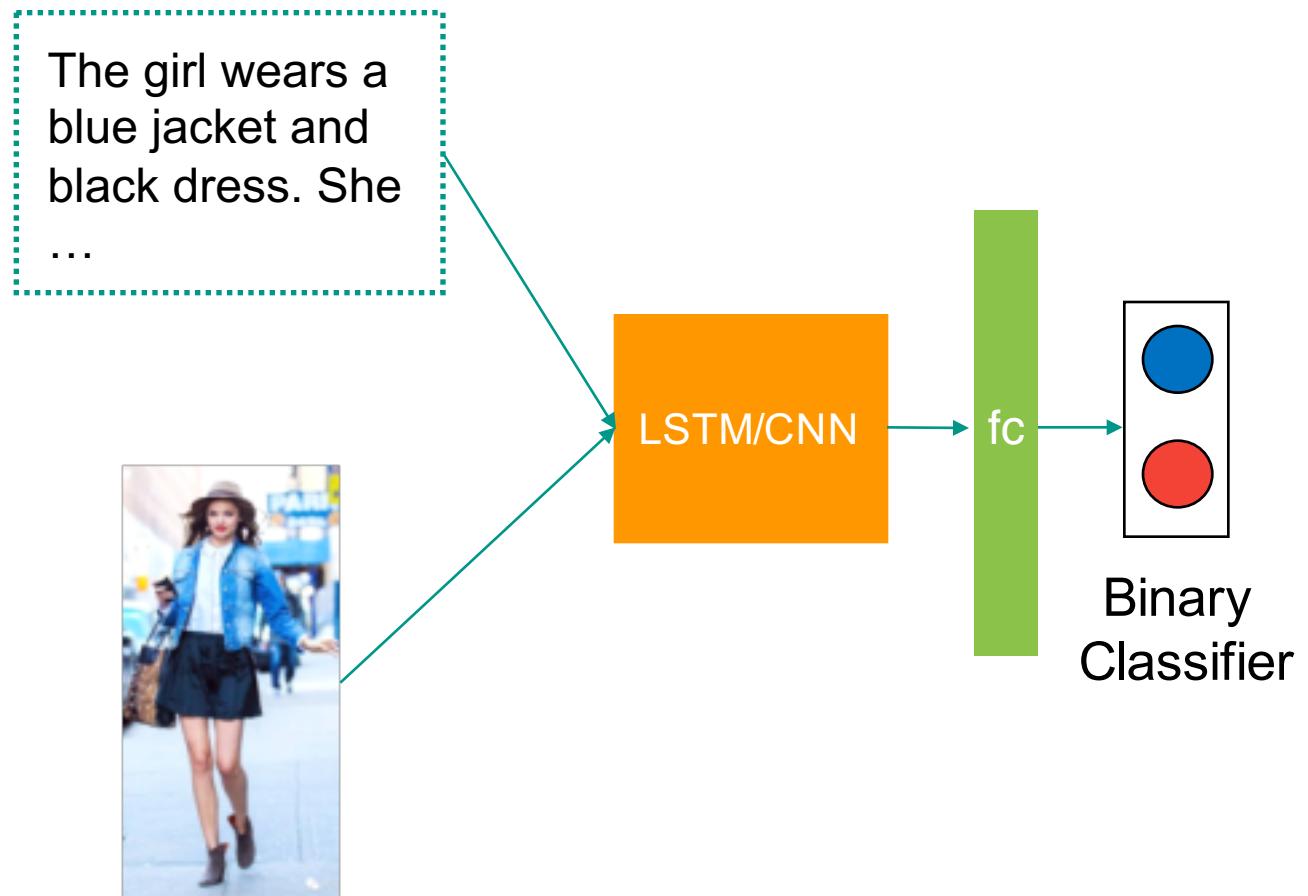
## Retrieval Results



Person Image Database

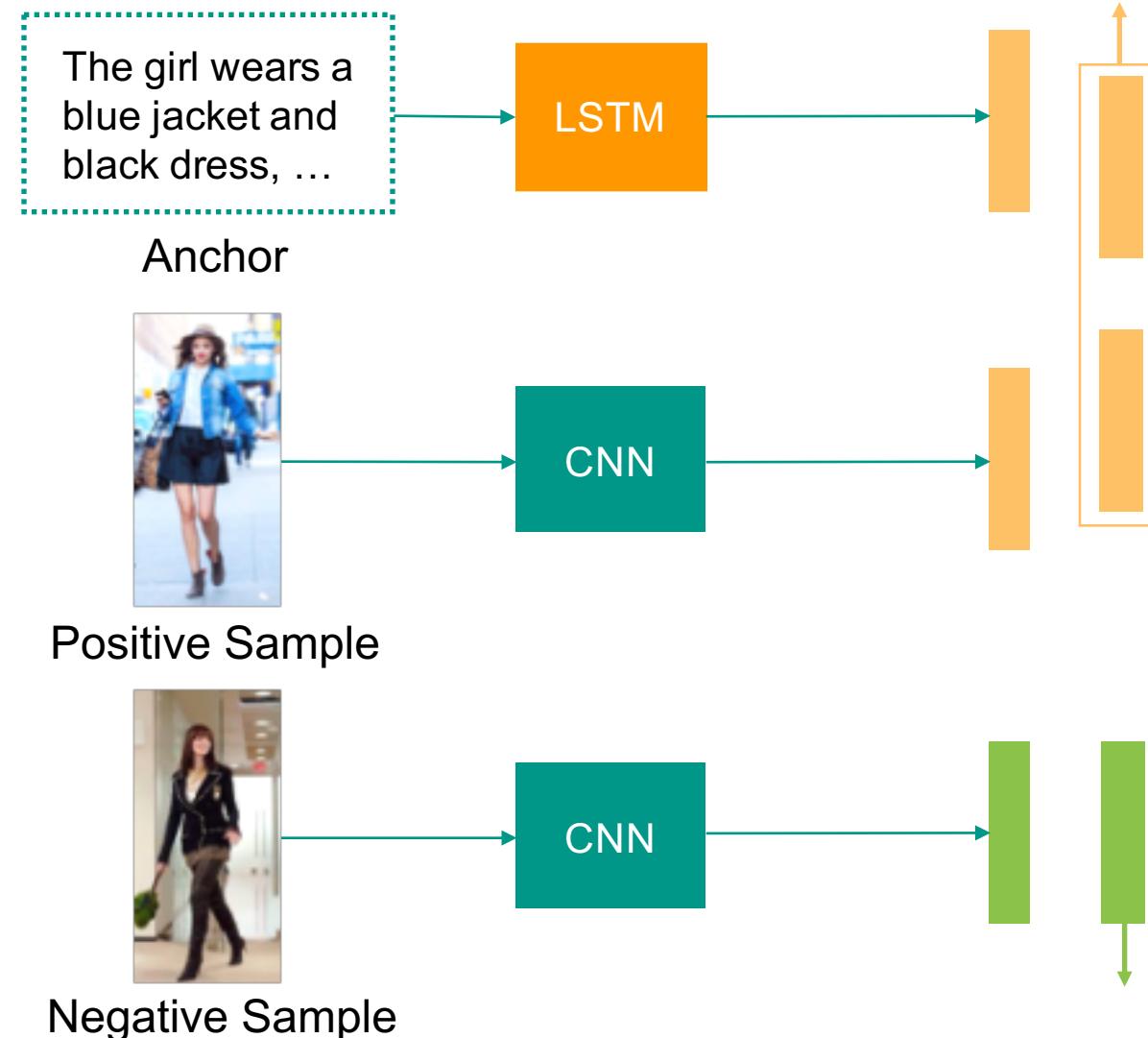
# Possible Solutions

- Pair-wise Loss
- Triplet Loss
- Classification Loss



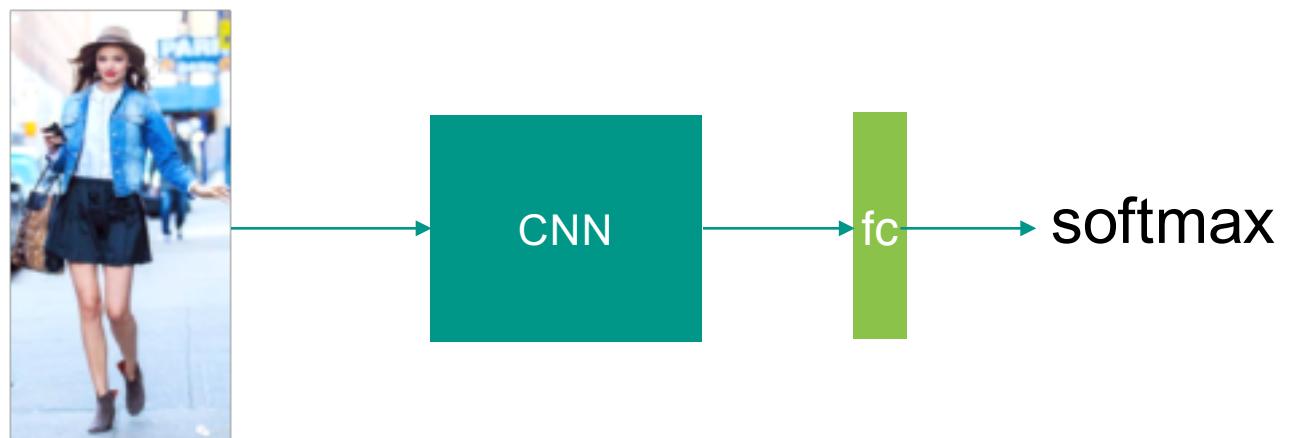
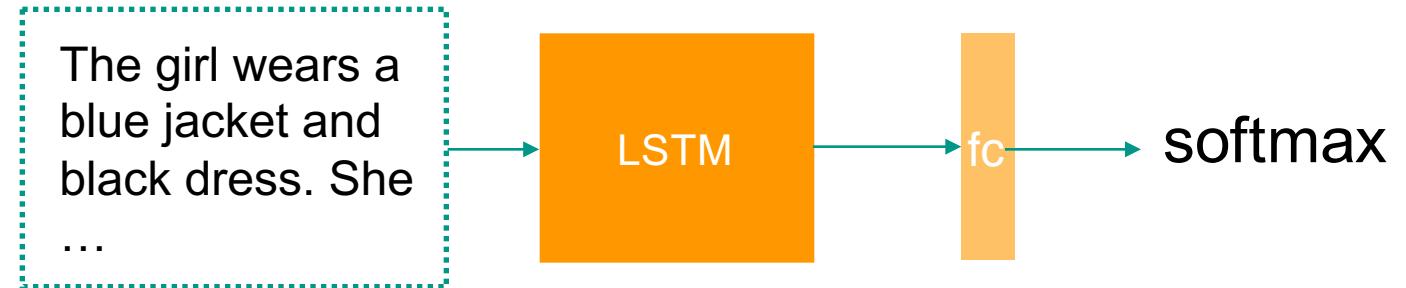
# Possible Solutions

- Pair-wise Loss
- Triplet Loss
- Classification Loss



# Possible Solutions

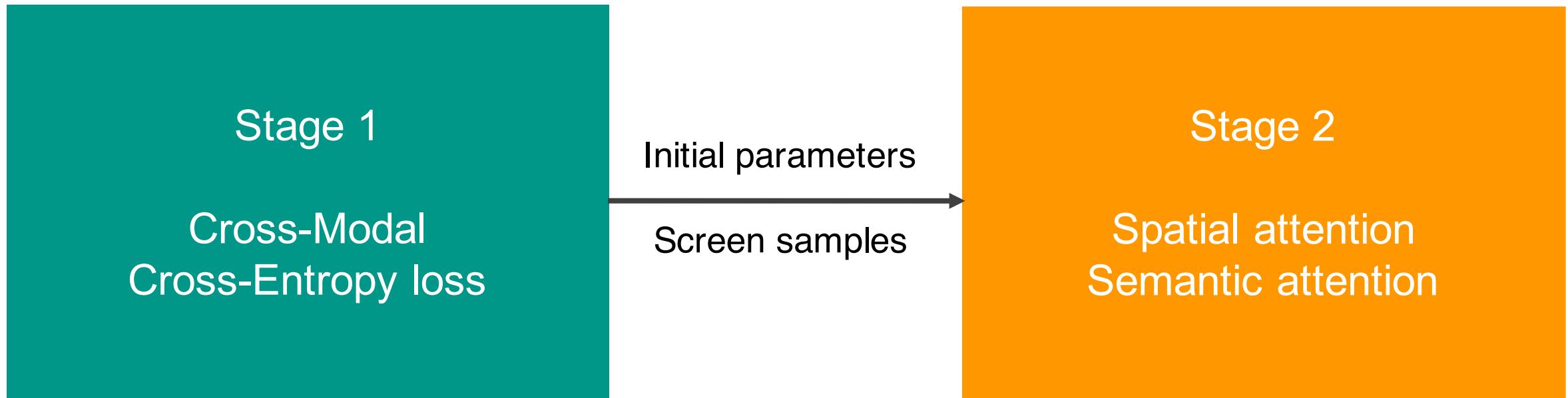
- Pair-wise Loss
- Triplet Loss
- Classification Loss



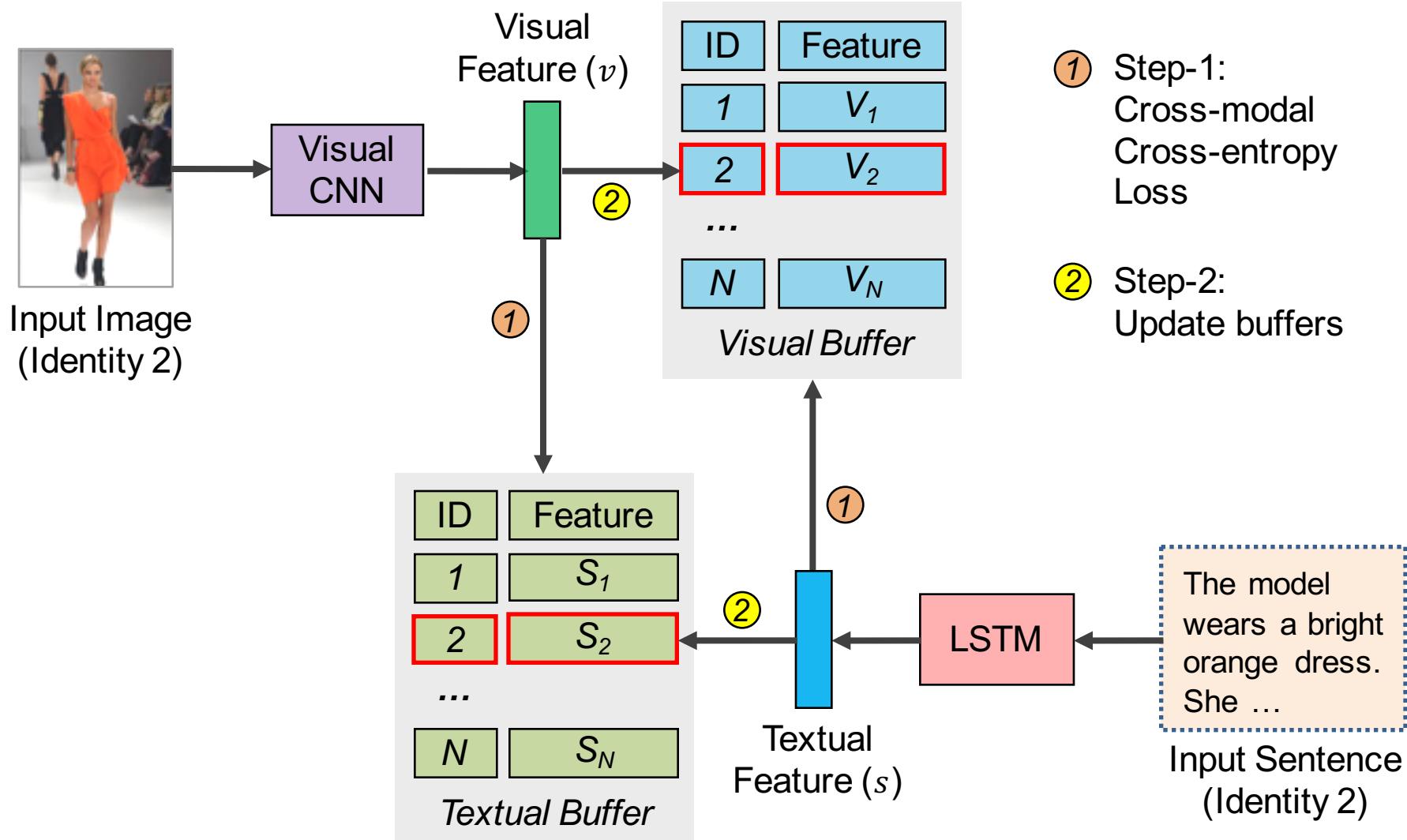
# Limitations of Possible Solutions

Solutions	Phase	Limitations
Pair-wise Loss	Test	$N^2$ visual-textual pairs
Triplet Loss	Train	Difficulty in selecting hard negative samples
Classification Loss	Train	Few positives for classifier matrix; different feature embedding space

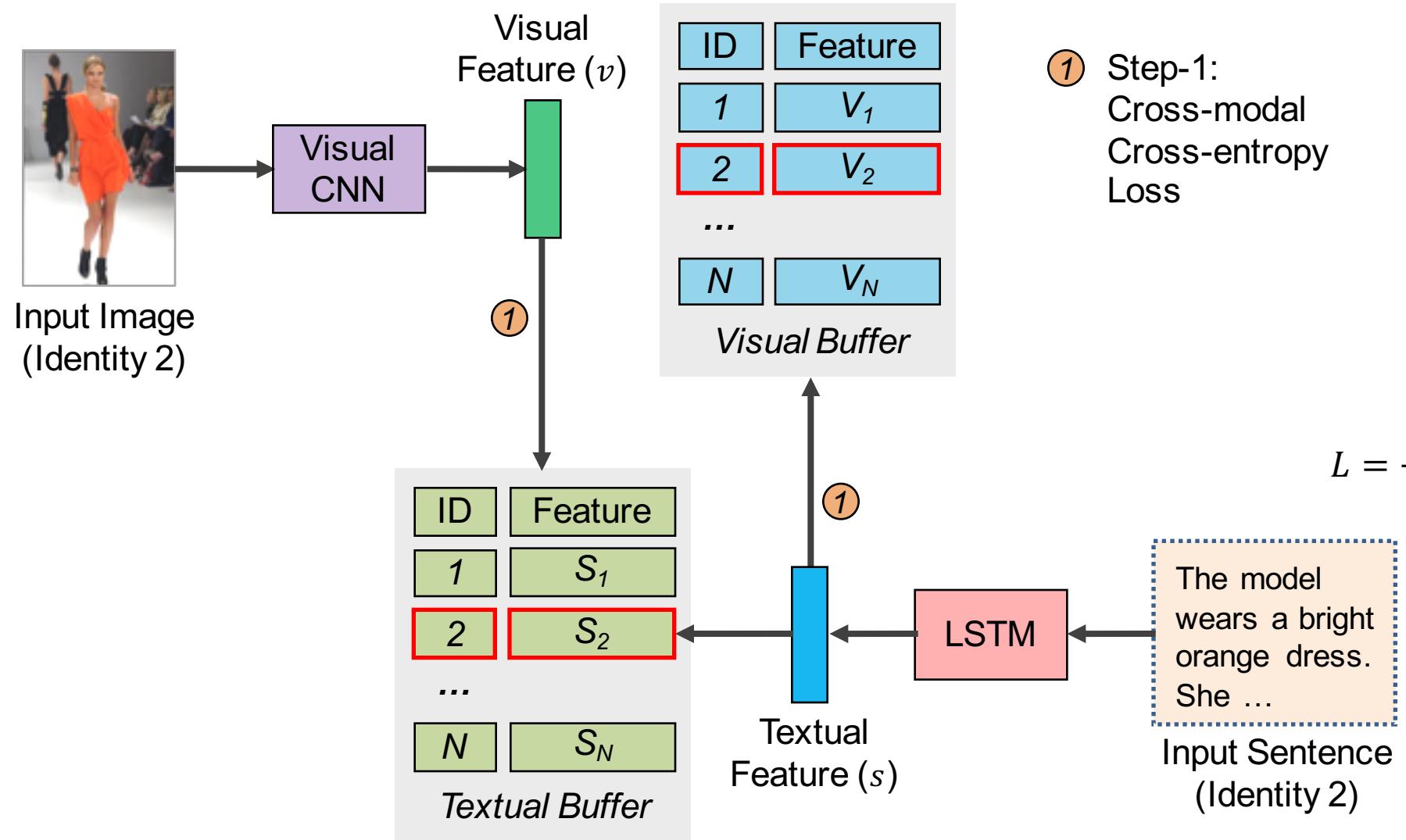
# Framework Overview



# Stage-1 CNN-LSTM with CMCE loss



# Stage-1 CNN-LSTM with CMCE loss



visual feature:  $v$

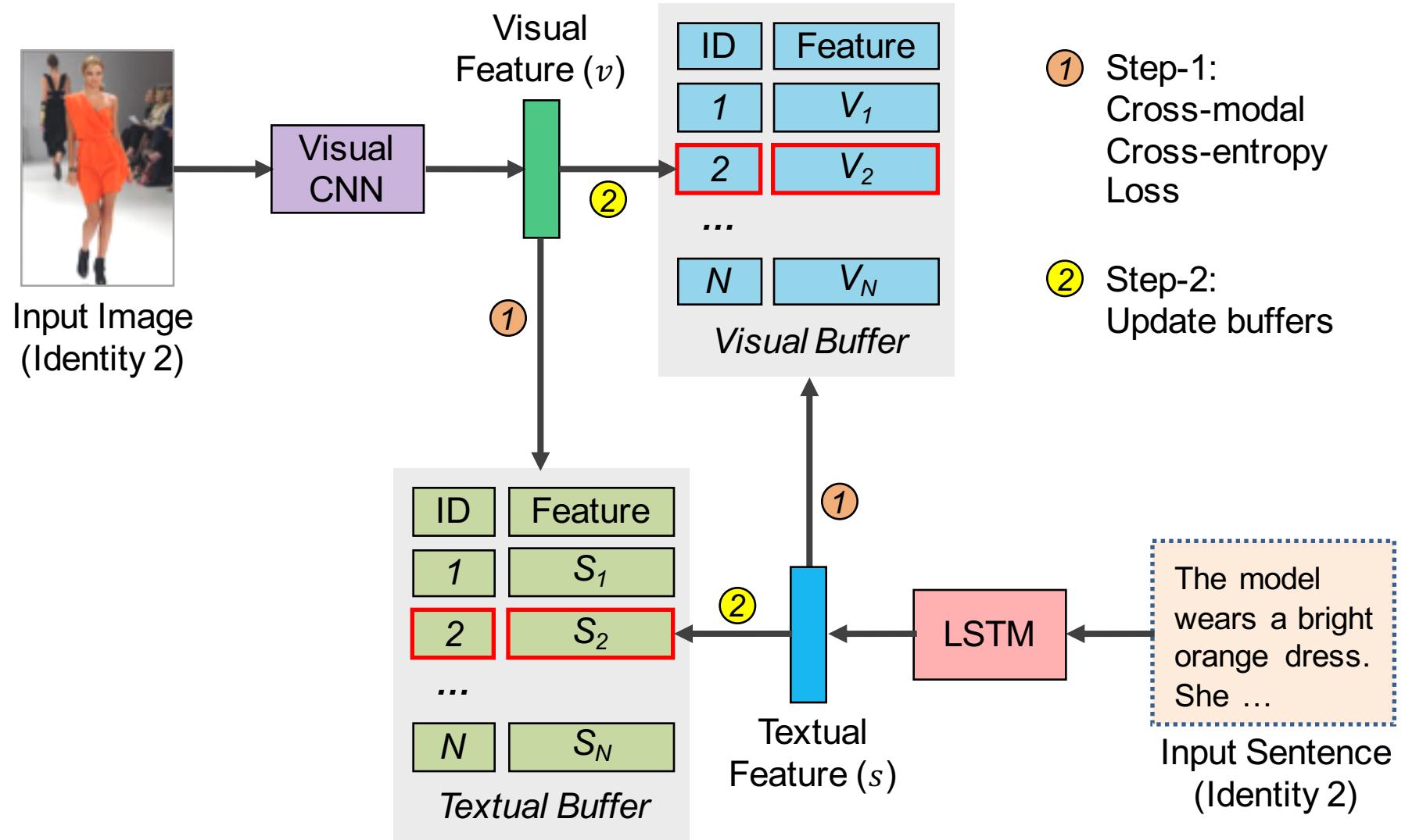
$$p_i^S(v) = \frac{\exp(S_i^T v)}{\sum_{j=1}^N \exp(S_j^T v)}$$

$$p_i^V(s) = \frac{\exp(V_i^T s)}{\sum_{j=1}^N \exp(V_j^T s)}$$

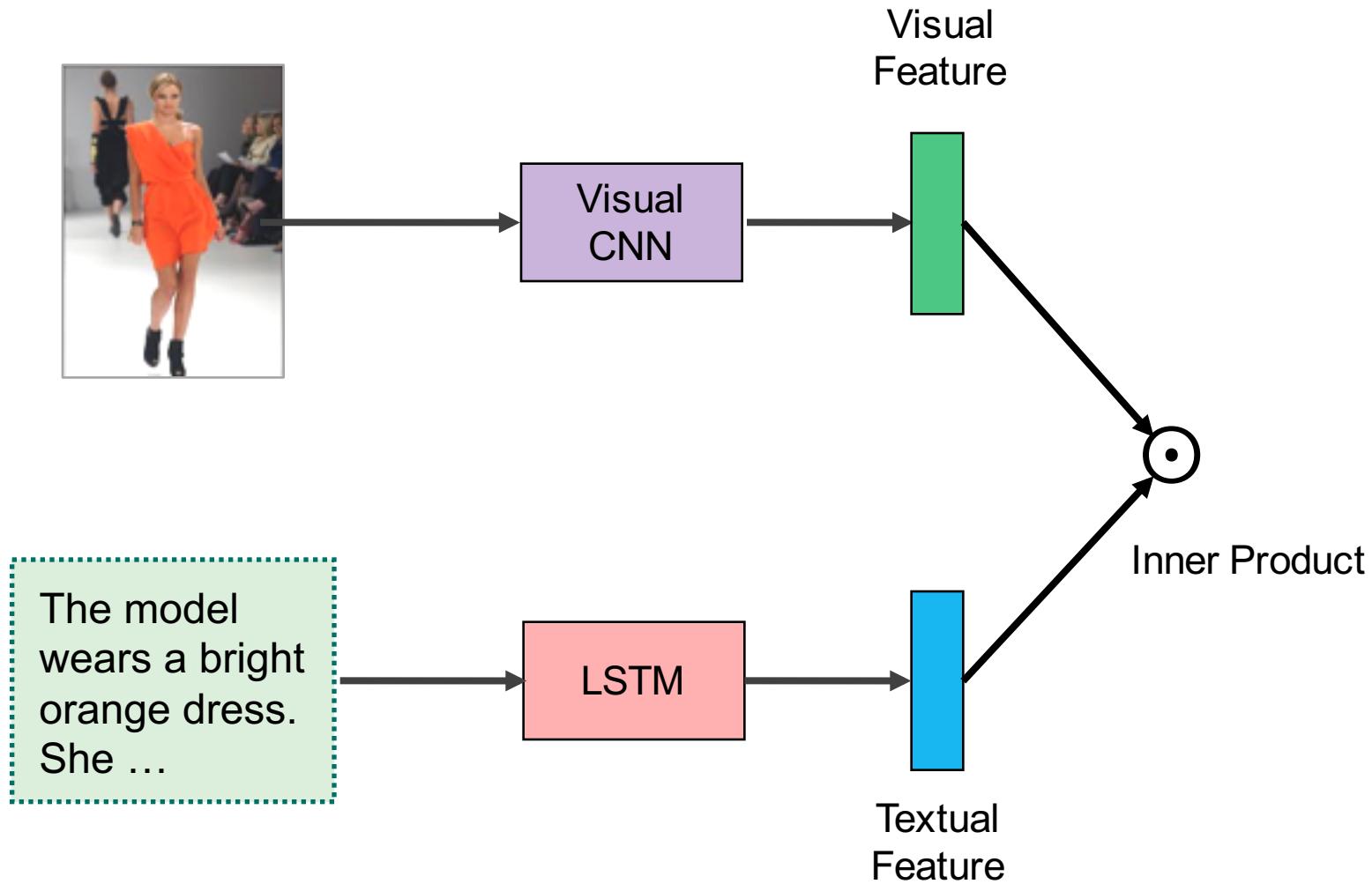
$$L = - \sum_v \log p_{t_v}^S(v) - \sum_s \log p_{t_s}^V(s)$$

The model wears a bright orange dress. She ...  
Input Sentence (Identity 2)

# Stage-1 CNN-LSTM with CMCE loss



# Stage-1 Test Phase



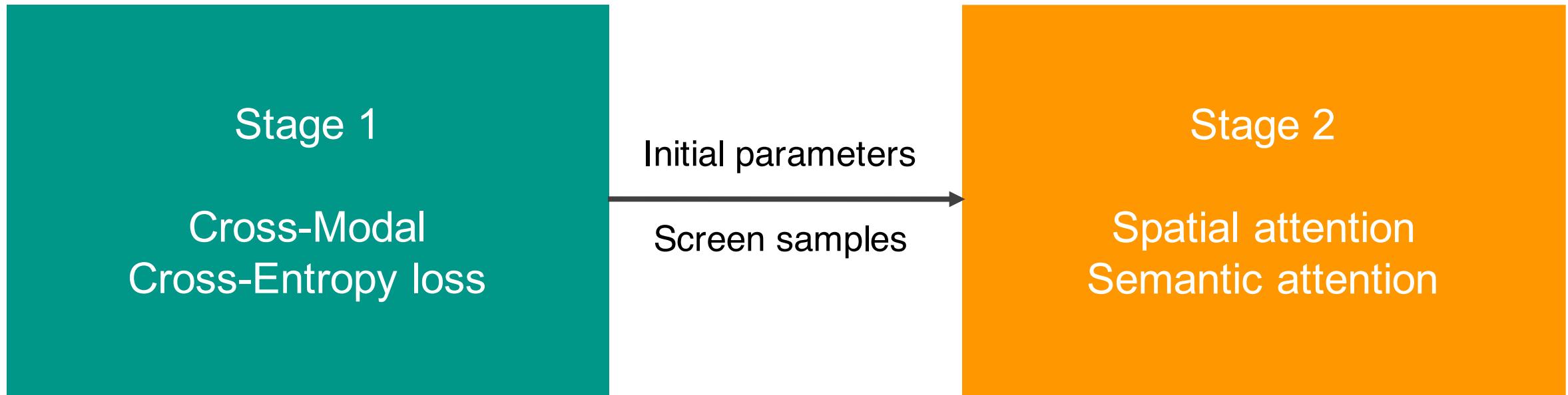
# CMCE vs. Others

Solutions	Phase	Limitations	CMCE
Pair-wise Loss	Test	$N^2$ visual-textual pairs	$N$ images + $N$ texts
Triplet Loss	Train	Difficulty in selecting hard negative samples	Hard negative data in each epoch
Classification Loss	Train	Few positives for classifier matrix; different feature embedding space	Non-parametric; Same feature subspace

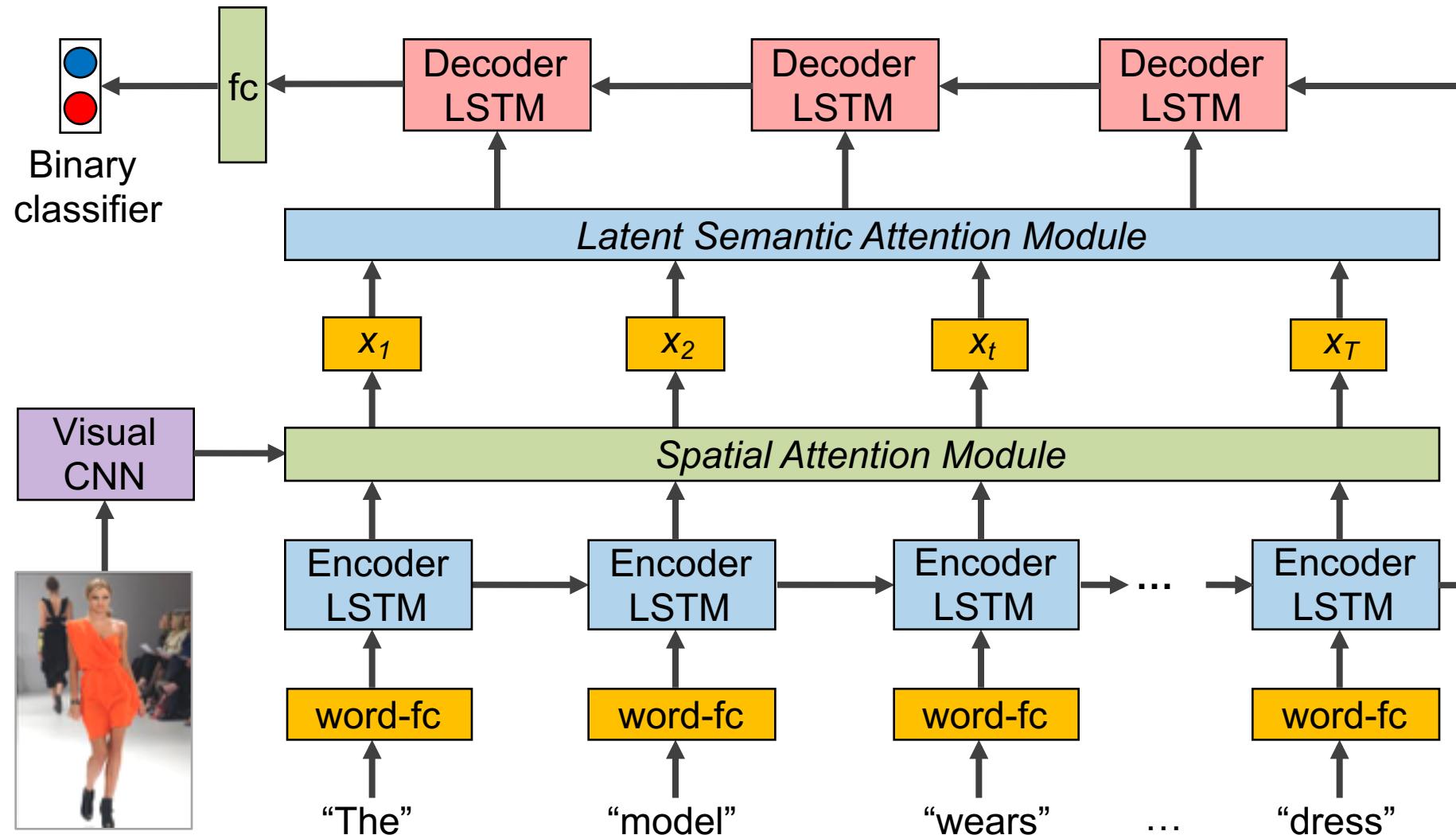
# Problems in Stage-1

- Compress the whole sentence into a single vector
- Highly focus on the latest words
- Sensitive to sentence structure variations
  - “The girl who has brown hair is wearing a **white dress**.”
  - “The girl wears a white dress. She has **brown hair**.”

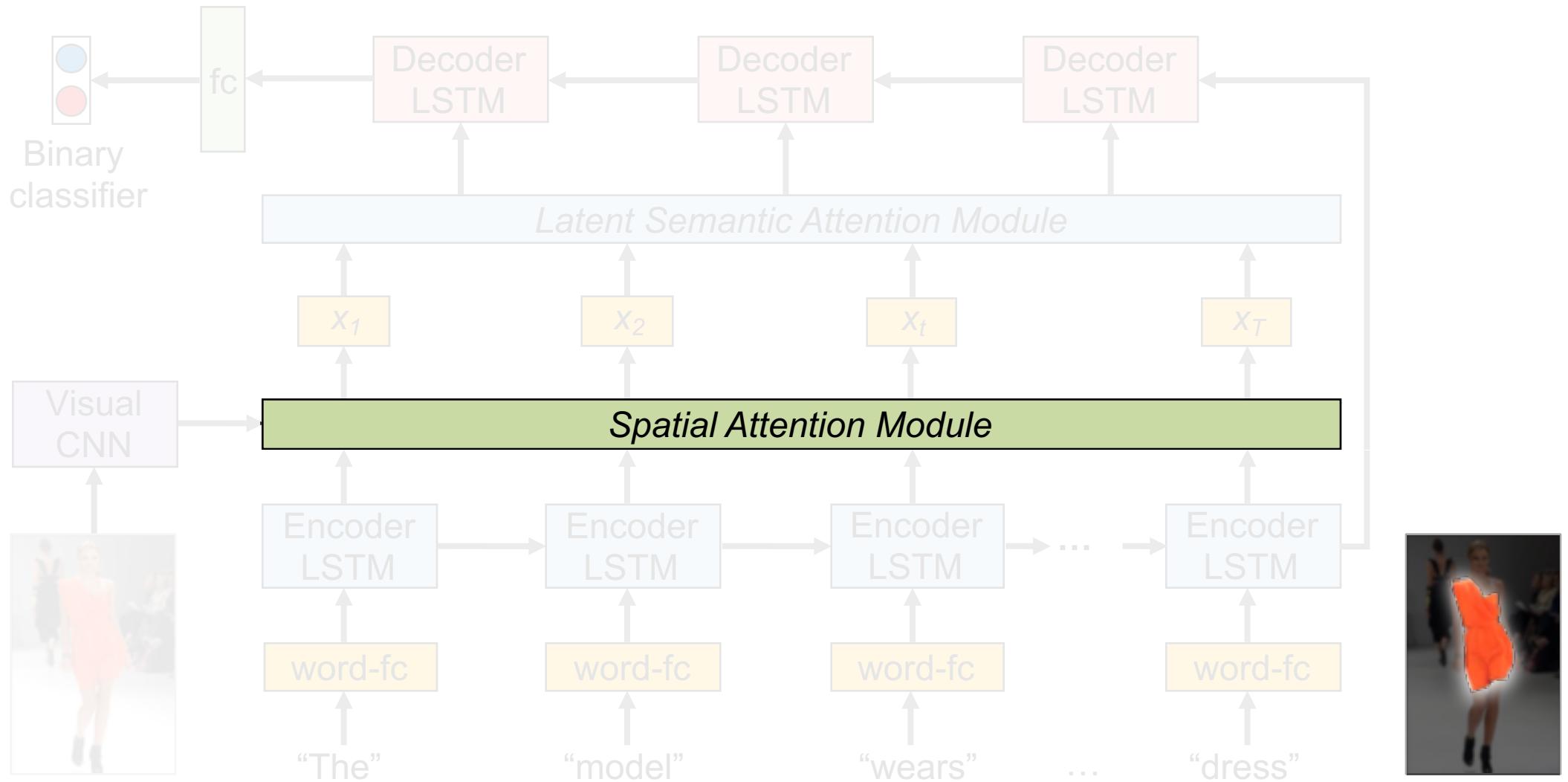
# Framework Overview



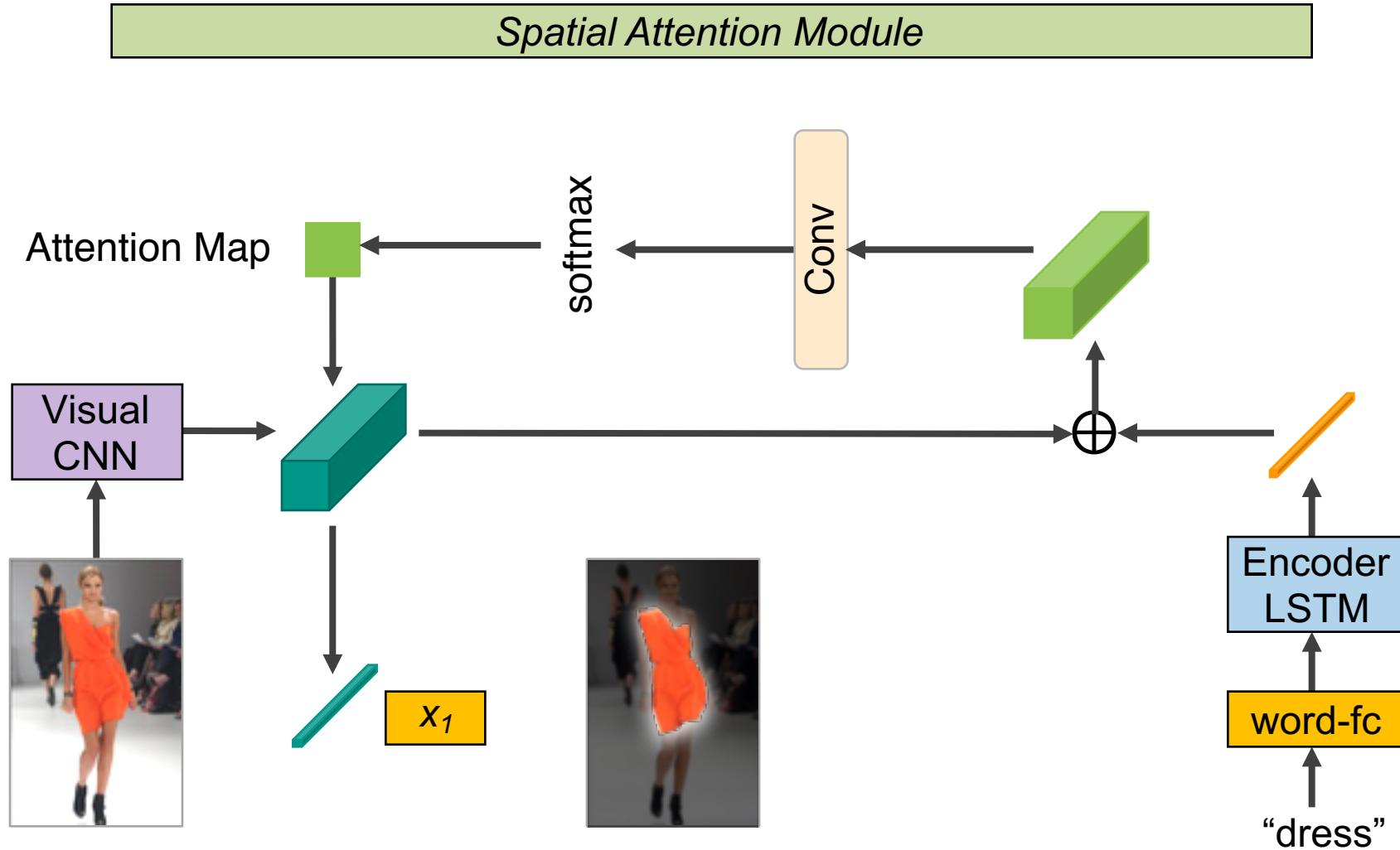
# Stage-2 CNN-LSTM with latent co-attention



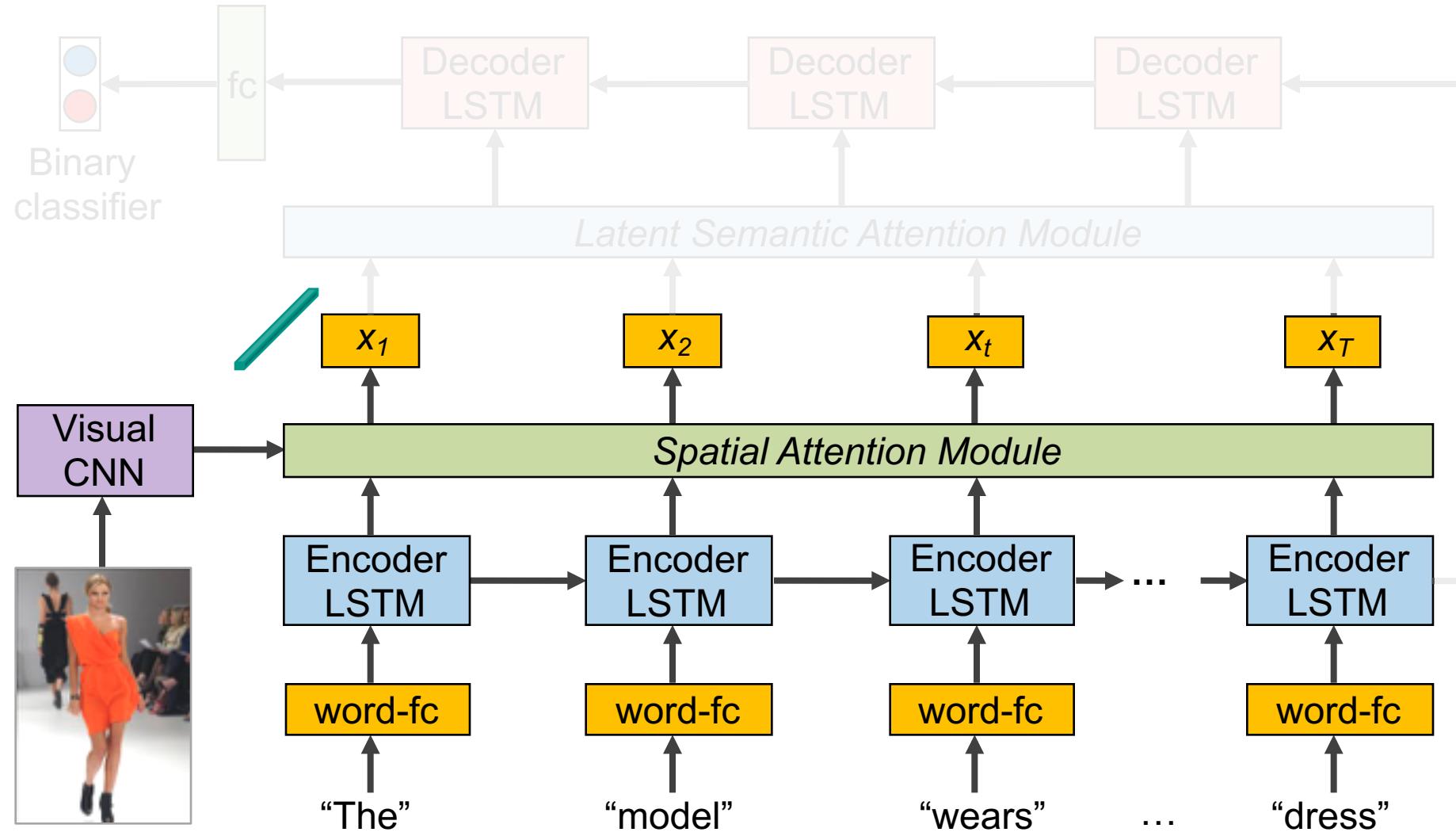
# Stage-2 CNN-LSTM with latent co-attention



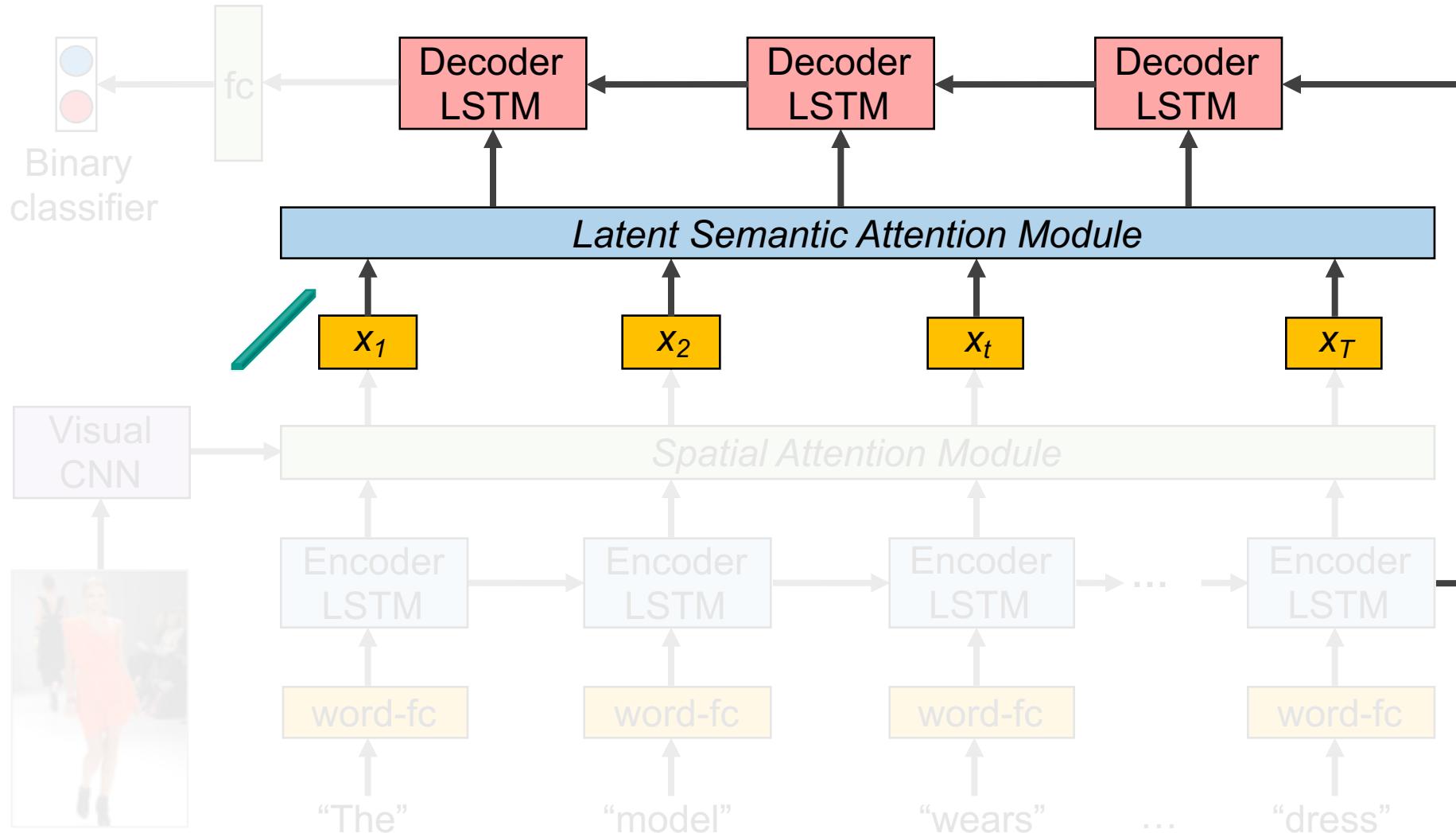
# Spatial Attention



# Spatial Attention



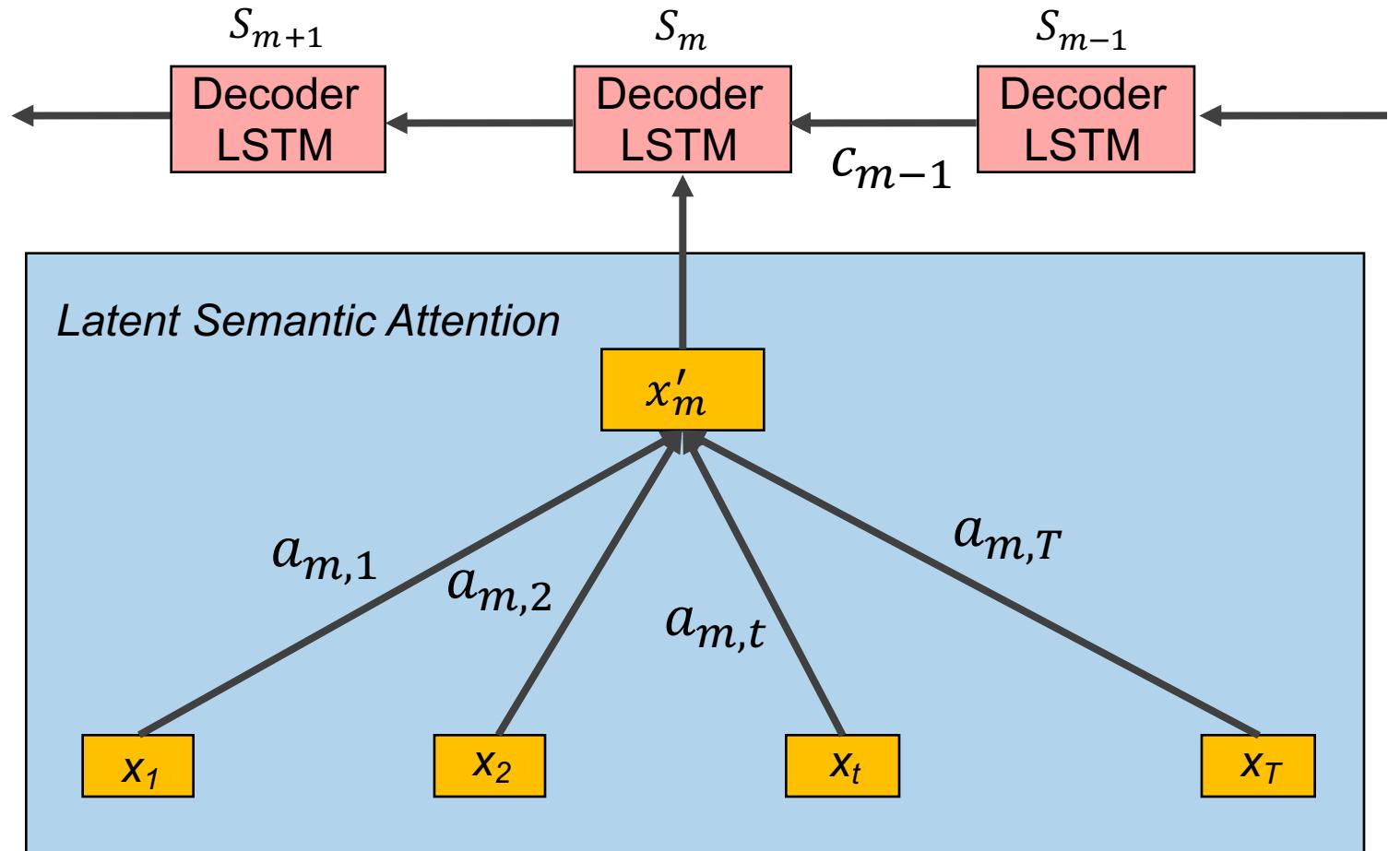
# Latent Semantic Attention



# Latent Semantic Attention

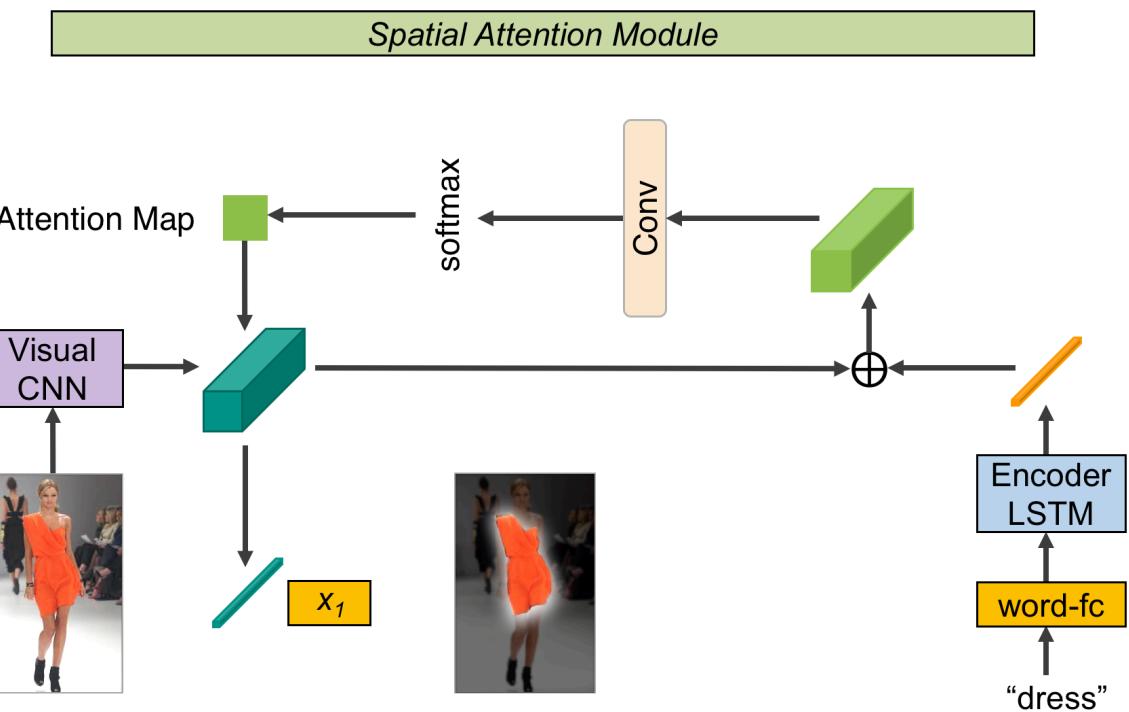
$$x'_m = \sum_{t=1}^T a_{m,t} x_t$$

$$a_{m,t} = \text{softmax}(f(c_{m-1}, x_t))$$



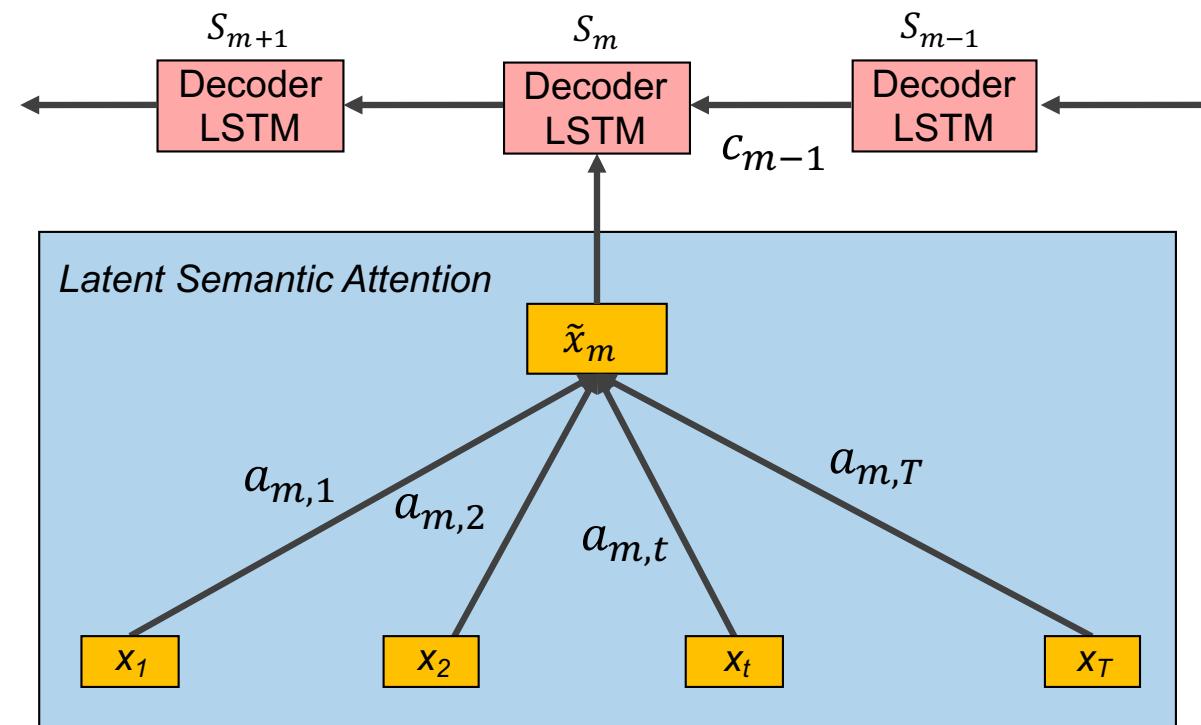
# Problems in Stage-1

- Compress the whole sentence into a single vector
- Highly focus on the latest words
- Sensitive to sentence structure variations



# Problems in Stage-1

- Compress the whole sentence into a single vector
- Highly focus on the latest words
- Sensitive to sentence structure variations



# Experiments

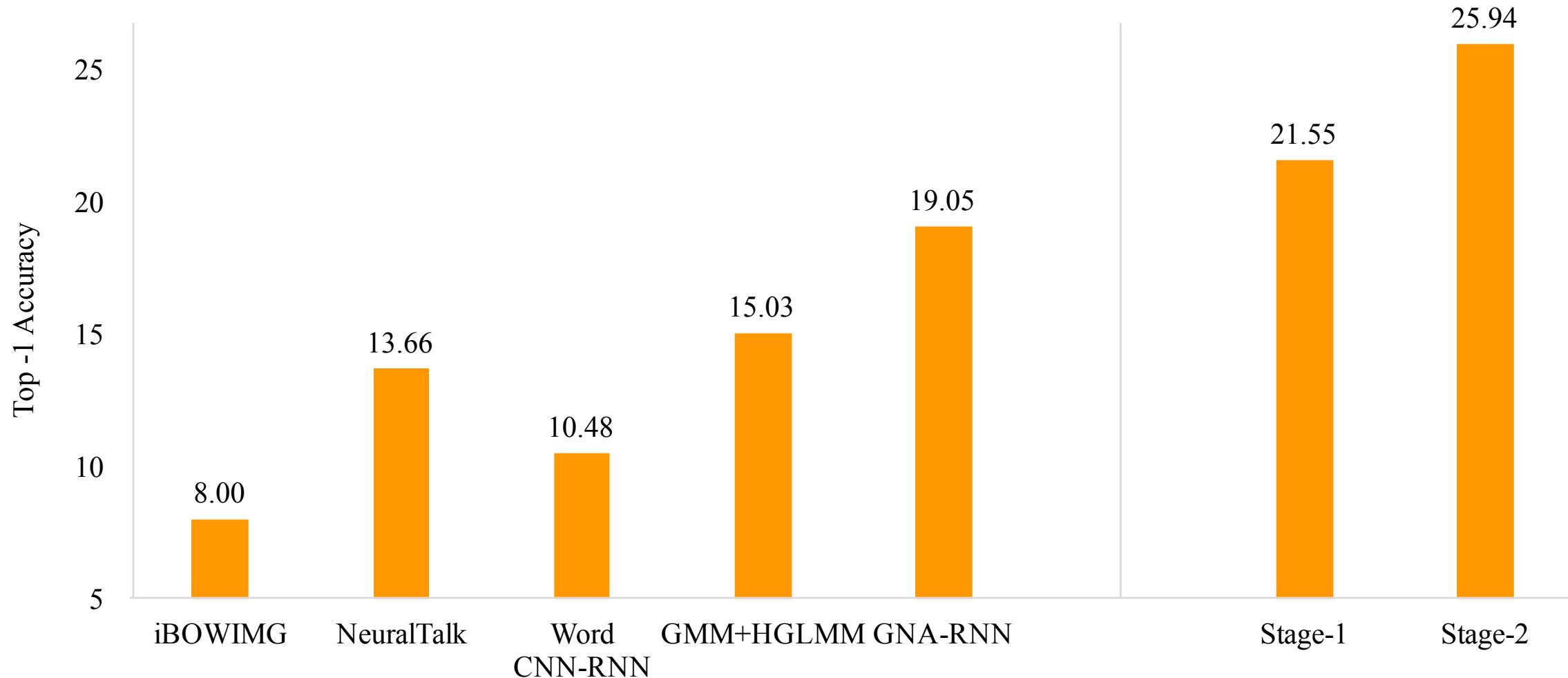
# CUHK-PEDES Dataset

13,003 Identities

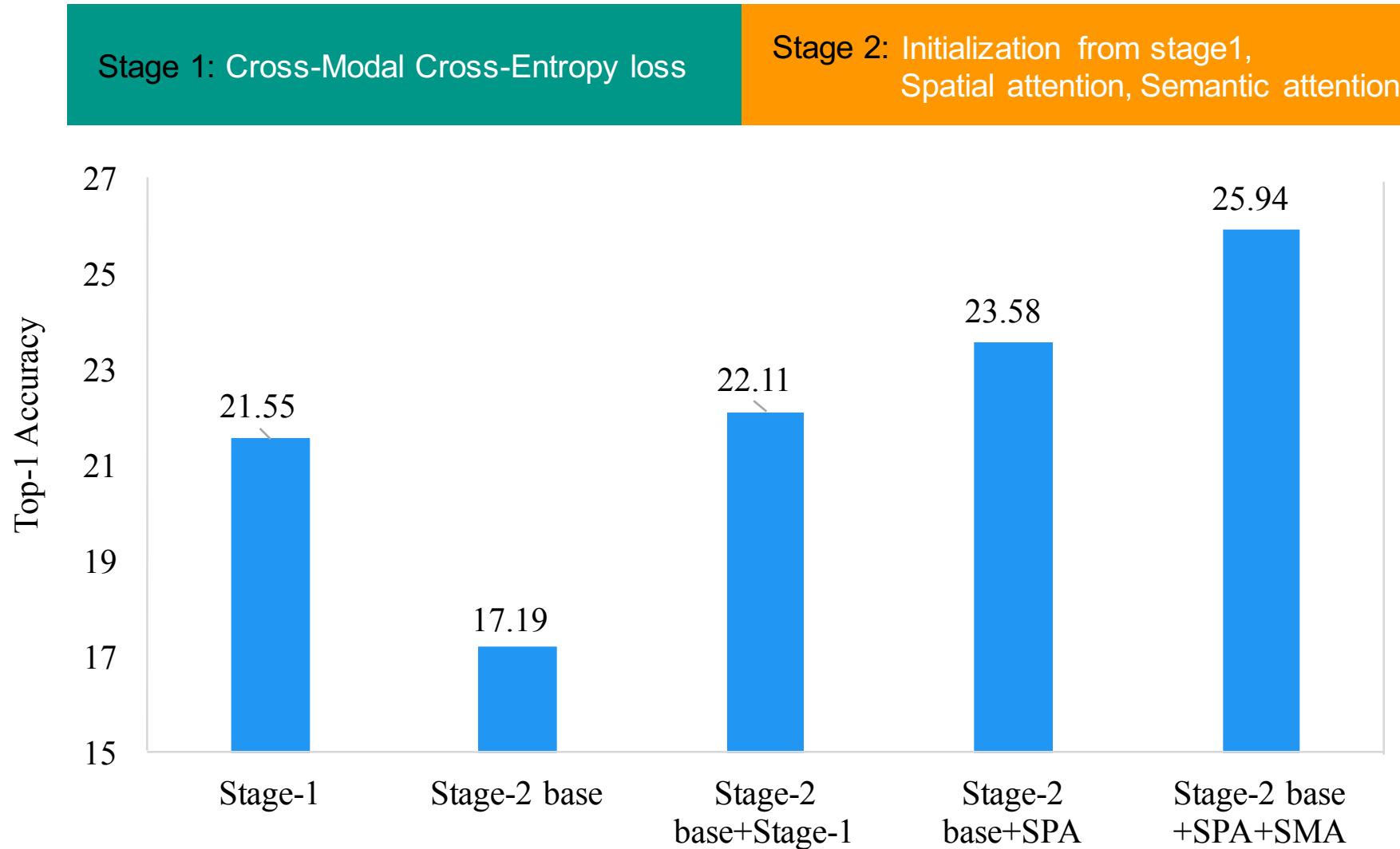
40,206 Images

80,412 Descriptions

# Comparison with different methods



# Performance Analysis



# Qualitative Results



The man is wearing a blue and white striped shirt and black pants. He is also wearing black shoes.



A man is wearing an orange button up shirt, a pair of white pants, and a pink umbrella in his right hand.

# Outcomes

- Propose a Cross-modal Cross-entropy loss function in Stage-1
- Present a Spatial attention and Semantic attention in Stage-2
- Collect a dataset for natural language based person search

# Future Work

# Future Work

- Combine social relationship with appearance cues
- Use language to assist visual feature learning

# Person Re-id Datasets



# Social Relationships



# Future Work

- Combine social relationship with appearance cues
- Use language to assist visual feature learning

# Textual-Visual Datasets

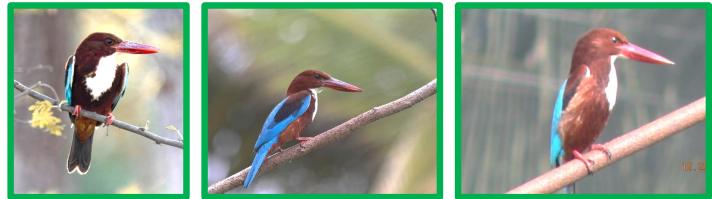
CUB Dataset



Flower Dataset



# Retrieval Results



A long red beaked bird with a white breast and turquoise secondaries.



This bird has a blue head and back with dark secondaries.



These petals are large and purple in color with green in the middle.



The petals on this flower are mostly white with a yellow stamen.

# Book/Movie Generator





# Thank You

Shuang Li  
The Chinese University of Hong Kong

