Person Search with Natural Language Description

Shuang Li, Tong Xiao, Hongsheng Li, Bolei Zhou, Dayu Yue, Xiaogang Wang
The Chinese University of Hong Kong, Massachusetts Institute of Technology, SenseTime Group Limited

Language-Based Person Search

Finding person images in image databases given the person's language descriptions.

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

Gated Neural Attention (GNA-RNN)

- Latent visual concepts (blue branch)

$$v = CNN(Img), v \in \mathbb{R}^d$$

- Word-concept attention

$$A_t = f_1\big(LSTM(W_t)\big), A_t \in \mathbb{R}^d$$

- Word-level importance

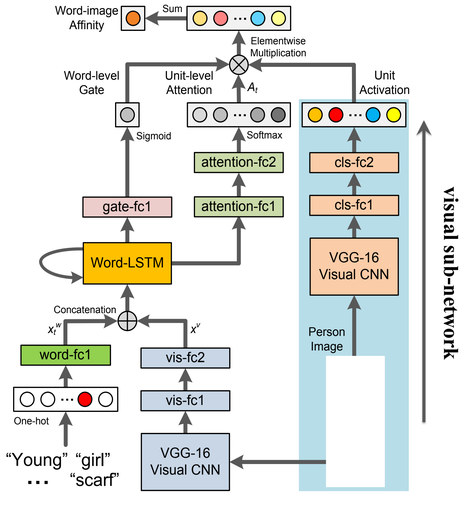
$$g_t = f_2\big(LSTM(W_t)\big), g_t \in \mathbb{R}$$

- Word-image affinity

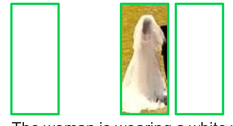
$$\hat{a}_t = g_t \sum_{n=1}^{d} A_t(n) v_n, \quad \sum_{n=1}^{d} A_t(n) = 1$$

- Sentence-image affinity

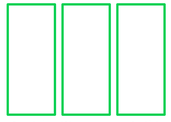
$$\hat{a} = \sum_{t=1}^{T} \hat{a}_t$$



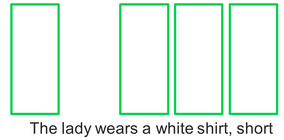
Qualitative Search Results



The woman is wearing a white wedding dress with brown hair pulled back into a long veil. The dress is cinched with a white ribbon belt.

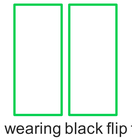


A girl with a ponytail is wearing a red top with gray denim thigh-high skirt. She is carrying a yellow shoulder bag and looking at a phone.

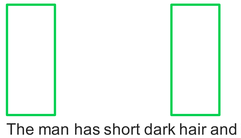


The lady wears a white shirt, short flowered skirt as she stands on the floor.

A man is wearing an orange polo shirt and black pants, and he is walking away.



He is wearing black flip flops, black shorts with a white stripe on the side, and a yellow shirt with short sleeves and a rounded neck.



The man has short dark hair and is wearing a white tuxedo jacket, white tuxedo shirt, black bow tie, black dress pants, and black shoes.

CUHK-PEDES Dataset

The woman is dressed up like Marilyn Monroe, with a white dress that is blowing upward in the wind, short curly blonde hair, and high heels.

The man is wearing blue scrubs with a white lab coat on top. He is holding paperwork in his hand and has a name badge on the left side of his coat.

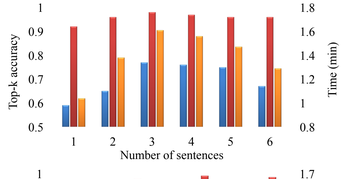
#dataset	#img	#person	#img/person
5 (re-id)	40206	13003	3.09

#sent.	#word	#sent./img	#word/sent.
80412	9408	2	23.5

User Study on CUHK-PEDES

- Expressive power in terms of the number of sentences and sentence length.
- Expressive power of different word types.

	orig. sent.	w/o nouns	w/o adjs	w/o verbs
top1	0.59	0.38	0.44	0.57
top5	0.92	0.81	0.85	0.92
time	1.14	1.01	0.98	1.12



Quantitative Results

	NeuralTalk [1]	CNN-RNN [2]	EmbBoW	GNA-RNN
top1	13.66	8.07	8.38	19.05
top10	41.72	32.47	30.76	53.64
	QAWord	QAWord-img	QABoW	-
top1	11.62	10.21	8.00	-
top10	42.42	44.53	30.56	-

	GNA-RNN	w/o pre-train	w/o gates	w/o attention
top1	19.05	8.93	13.86	4.85
top10	53.64	32.32	44.27	27.16

[1] Show and tell: A neural image caption generator, *CVPR*, 2015
[2] Learning deep representations of fine-grained visual descriptions, *CVPR*, 2016