Visual-Semantic Alignment Across Domains Using a Semi-Supervised Approach

A. Carraggi, M. Cornia, L. Baraldi, R. Cucchiara

AImage Lab

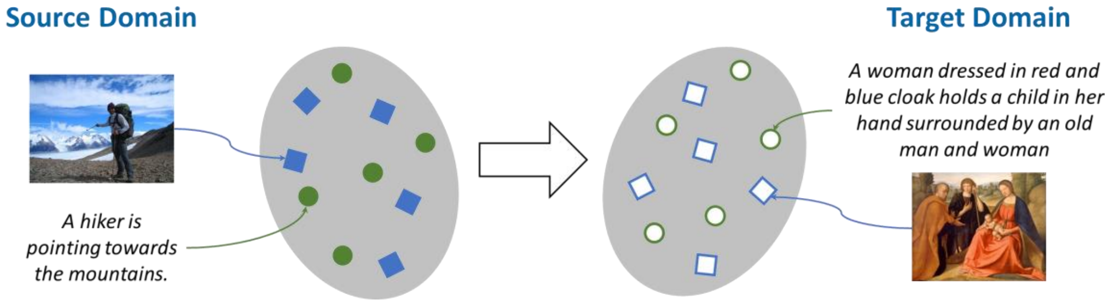
UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

INTRODUCTION

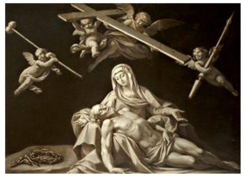
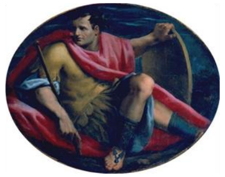
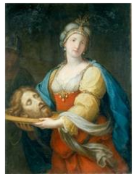
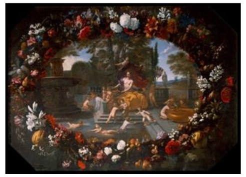
Visual-semantic embeddings have been extensively used as a powerful model for cross-modal retrieval of images and sentences.

Contributions:

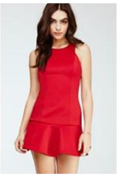
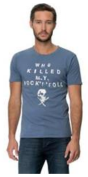
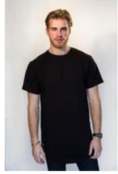
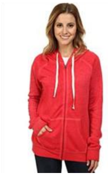
- A novel domain adaptation model for cross-modal retrieval, in which the knowledge learned from a supervised dataset can be transferred on a target dataset.
- Application to fashion and cultural heritage datasets.


Source Domain
Target Domain
A hiker is pointing towards the mountains.
A woman dressed in red and blue cloak holds a child in her hand surrounded by an old man and woman

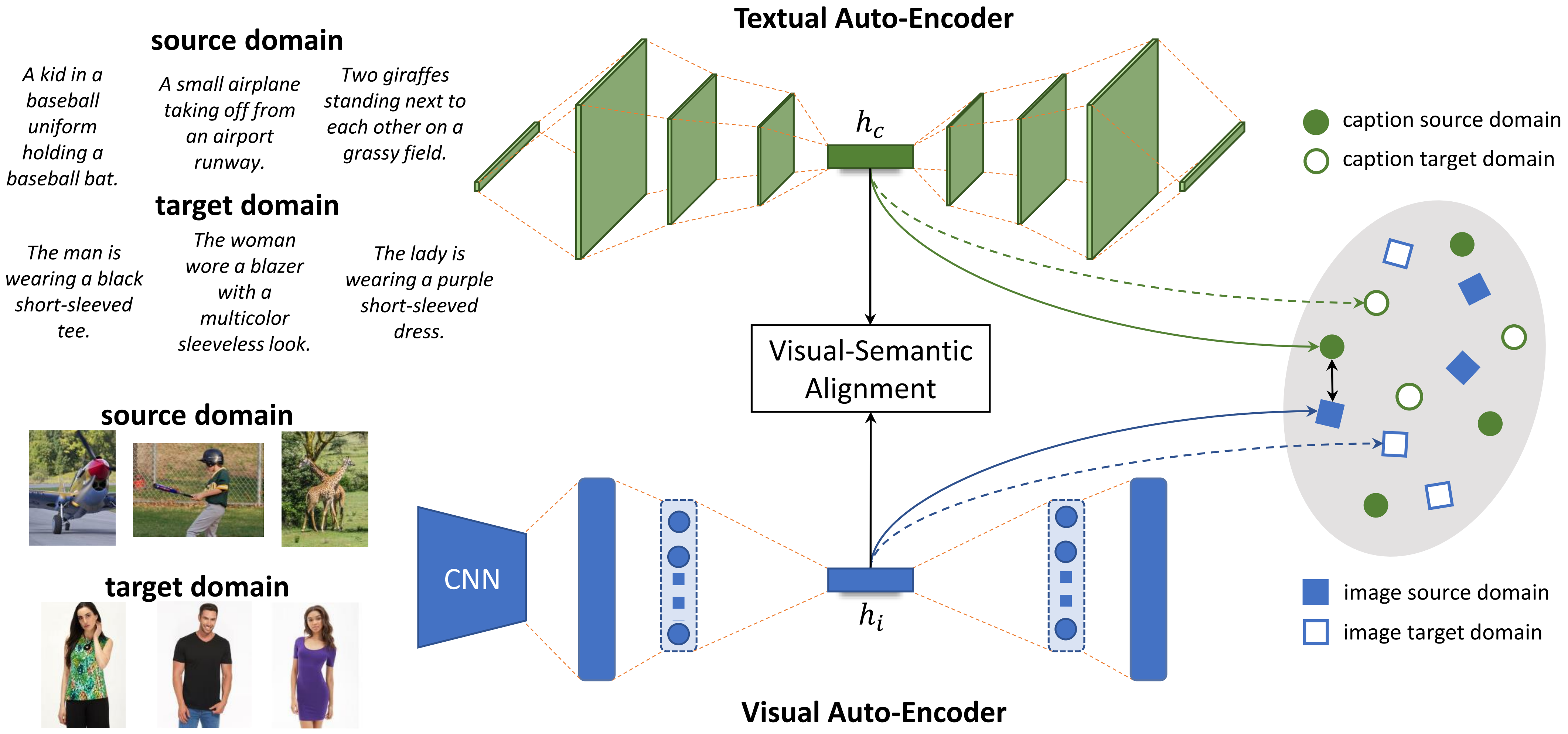
Cultural Heritage Domain

Query Caption	Top-1 Retrieved Image	Query Image	Top-1 Retrieved Caption
A woman holding a dead body with three angels flying over them.			*Armed man wearing a red cloak and holding a stick and a shield.*
A young lady with gold, red and blue dress is carrying a dead man's head on a plate.			*Woman and children in front of a fountain inside an oval frame made of flowers.*

Fashion Domain

Query Caption	Top-1 Retrieved Image	Query Image	Top-1 Retrieved Caption
The lady wore a red sleeveless dress.			*The man is wearing a blue short-sleeved tee printed with letters.*
The man is wearing a black short-sleeved tee.			*The lady is wearing a red long-sleeved hoodie.*

PROPOSED MODEL



Textual Auto-Encoder

source domain

A kid in a baseball uniform holding a baseball bat.

A small airplane taking off from an airport runway.

Two giraffes standing next to each other on a grassy field.

target domain

The man is wearing a black short-sleeved tee.

The woman wore a blazer with a multicolor sleeveless look.

The lady is wearing a purple short-sleeved dress.

source domain

target domain

Visual Auto-Encoder

● caption source domain

○ caption target domain

■ image source domain

□ image target domain

DATASETS

DeepFashion Dataset

- Fashion domain
- 78,979 images annotated with corresponding captions

A lady wears a gray short-sleeved tee printed with a skull.

The lady wore a red sleeveless dress.

The man is wearing a black long-sleeved jacket.

EsteArtworks Dataset

- 553 artworks from the Estense Gallery of Modena
- We collect textual sentences describing only the visual content of the artworks
- 1,278 image-sentence pairs

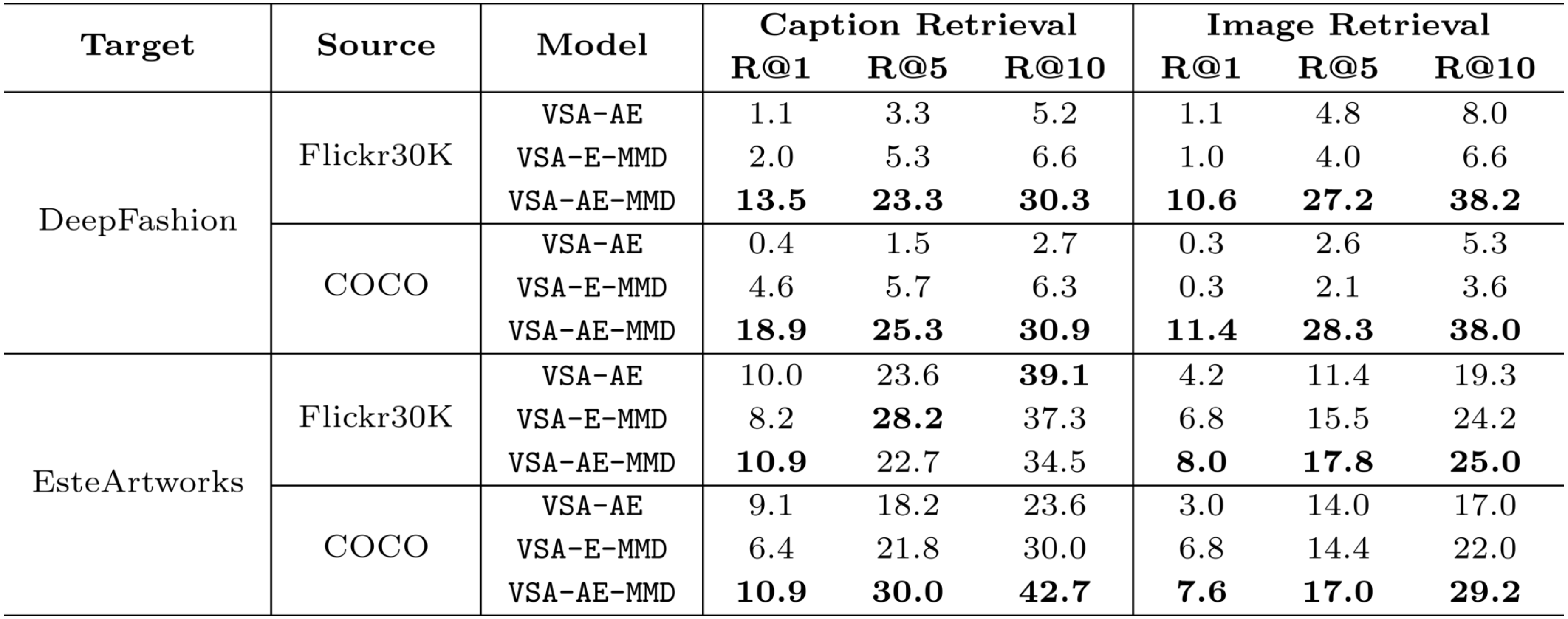
A young man with shoulder length hair, who is wearing a dark jacket with a white collar.

A man who is wearing a golden armour is sitting on a throne and is surrounded by a large group of people.

A man with a light-blue cloak and a dog are looking to an angel over them.

EXPERIMENTAL RESULTS

- **VSA-AE**: our model without the distribution alignment
- **VSA-E-MMD**: our model without the textual and visual reconstruction losses
- **VSA-AE-MMD**: our model with all its components



Target	Source	Model	Caption Retrieval			Image Retrieval		
			R@1	R@5	R@10	R@1	R@5	R@10
DeepFashion	Flickr30K	VSA-AE	1.1	3.3	5.2	1.1	4.8	8.0
		VSA-E-MMD	2.0	5.3	6.6	1.0	4.0	6.6
		VSA-AE-MMD	**13.5**	**23.3**	**30.3**	**10.6**	**27.2**	**38.2**
	COCO	VSA-AE	0.4	1.5	2.7	0.3	2.6	5.3
		VSA-E-MMD	4.6	5.7	6.3	0.3	2.1	3.6
		VSA-AE-MMD	**18.9**	**25.3**	**30.9**	**11.4**	**28.3**	**38.0**
EsteArtworks	Flickr30K	VSA-AE	10.0	23.6	**39.1**	4.2	11.4	19.3
		VSA-E-MMD	8.2	**28.2**	37.3	6.8	15.5	24.2
		VSA-AE-MMD	**10.9**	22.7	34.5	**8.0**	**17.8**	**25.0**
	COCO	VSA-AE	9.1	18.2	23.6	3.0	14.0	17.0
		VSA-E-MMD	6.4	21.8	30.0	6.8	14.4	22.0
		VSA-AE-MMD	**10.9**	**30.0**	**42.7**	**7.6**	**17.0**	**29.2**

Thank you!

marcella.cornia@unimore.it
aimagelab.ing.unimore.it

Angelo Carraggi

Marcella Cornia

Lorenzo Baraldi

Rita Cucchiara

Facebook AI Research has selected AImageLab as one of the 15 world-class research labs in Europe

Thanks to the **"CultMedia" Project** of the National Technology Cluster on Smart Communities, cofounded by the Italian Ministry of Education, University and Research (MIUR).