

Fashion2Events: A clothes recommendation method for social events

Jacopo Bartoli Jason Ravagli

Università degli Studi di Firenze
School of Engineering

17/06/2021



UNIVERSITÀ
DEGLI STUDI
FIRENZE





Outline

1. Introduction and Project Goal
2. Datasets
3. Proposed Method
4. Experiments
5. Application Example
6. Conclusions

Introduction

- Deep Learning have been applied with great results to the fashion field
 - Progress in CV techniques
 - High availability of data
- Fashion companies see in these emerging methods new opportunities to attract the customer
- Shift from academic research works to commercial applications
 - Retrieval of similar clothes and recommendation systems
 - Analysis and prediction of fashion trends
 - Automatic generation of new clothes from existing ones



Introduction - Fashion2Events

Project Goal

Study a recommendation method based on deep learning techniques that suggests at which type of social event it is most appropriate to wear a certain garment

Decomposition of the problem into **two separated tasks**:

- Detection and instance segmentation of clothes in an image
- Classification of the social event from the image of a single isolated garment



Datasets

Two different datasets, one for each task:

- **DeepFashion2** for the detection and segmentation of clothes
- **USED** for the social events classification



Datasets - DeepFashion2

- Published in 2019 by Ge et al. with the aim to provide a unified benchmark for clothes detection, segmentation, retrieval and landmark prediction
- About **491K images** of clothes belonging to **13 classes**
- Large variation in style, pose, colour, scale, occlusion and viewing angle between items



Datasets - DeepFashion2

- It comes already split into train/validation/test:
 - **Train set:** 391K (only **192K publicly available**)
 - **Validation set:** **34K**
 - **Test set:** 67K (labels **not publicly available**)
- We used the official **validation set as the test set**
 - The authors provided their best results on it to use as a comparison
- Further split the train set to use about 15% of the images (29K) as our validation set



Datasets - DeepFashion2



Datasets - USED

- About **525K images** of people attending to **14 different types of social events** (e.g. graduation, wedding, sport event)
 - Train set: 361K
 - Test set: 164K
- Large variations of image quality, illumination, places and number of people



Datasets - USED



Class: Graduation



Class: Picnic



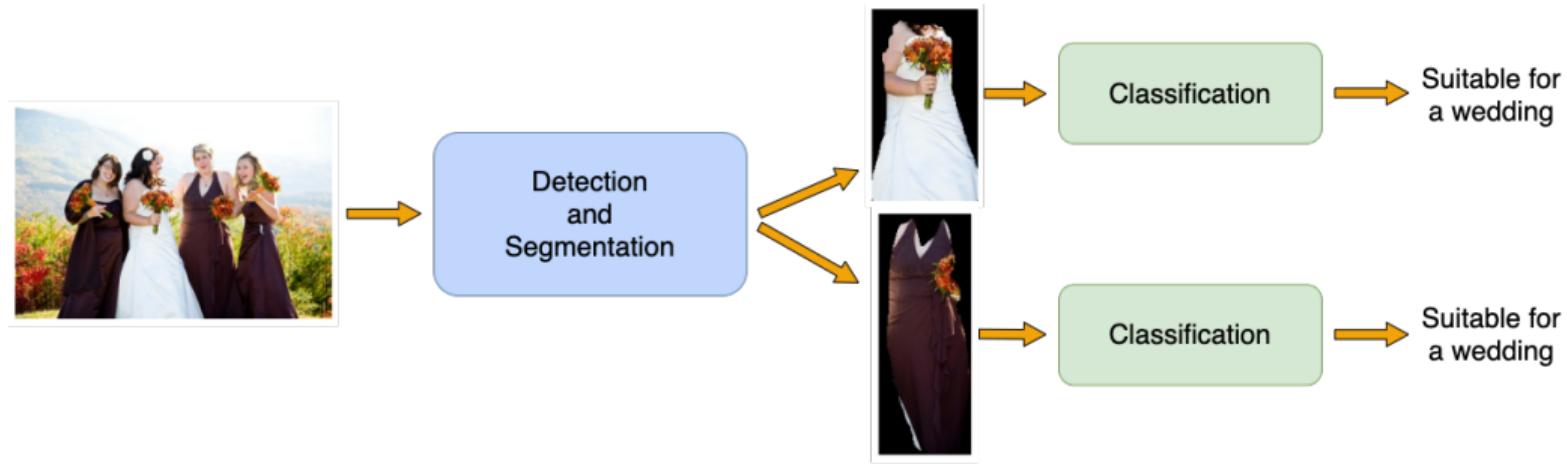
Class: Sport



Class: Concert



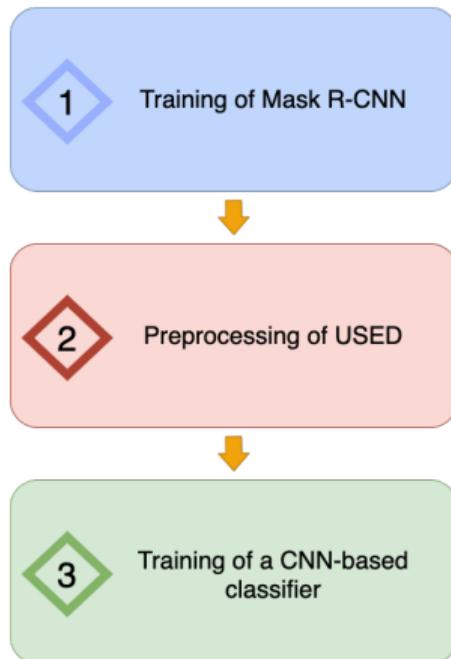
Proposed Method



Proposed Method - Training Pipeline

To build the recommendation system we performed **3 main steps**:

- Training of a detector/segmentator model (Mask R-CNN)
- Building of a new dataset made of images of single clothes using Mask R-CNN and USED
- Training of a CNN-based classifier of social events on the new dataset



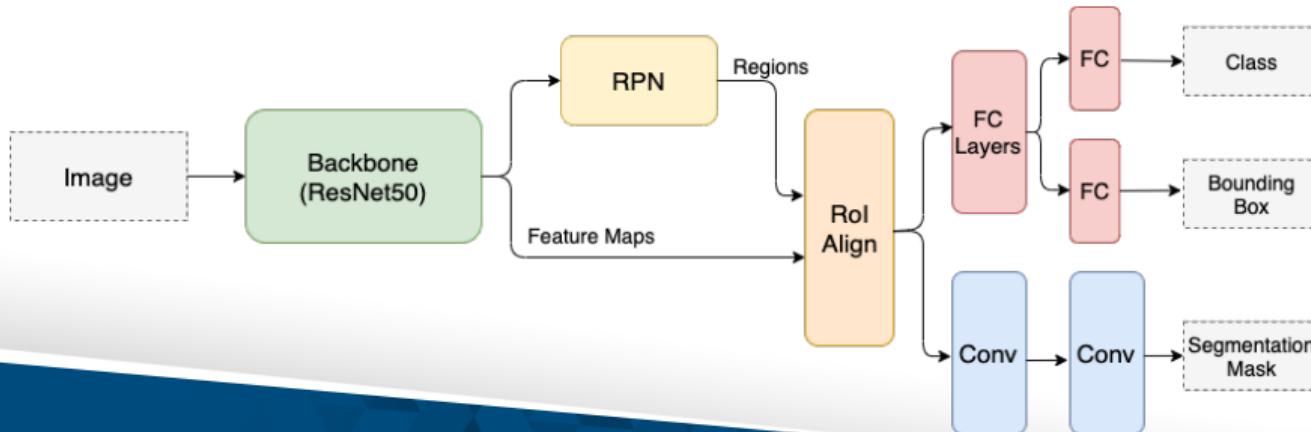
Step 1 - Mask R-CNN

- The authors of DeepFashion2 also proposed a deep learning model for fashion tasks: **Match R-CNN**
 - It solves the detection, segmentation, landmark estimation and retrieval tasks
 - Currently the best performing method on DeepFashion2
 - **An adaptation of Mask R-CNN to include retrieval**
- For detection and segmentation only the Mask R-CNN part is used



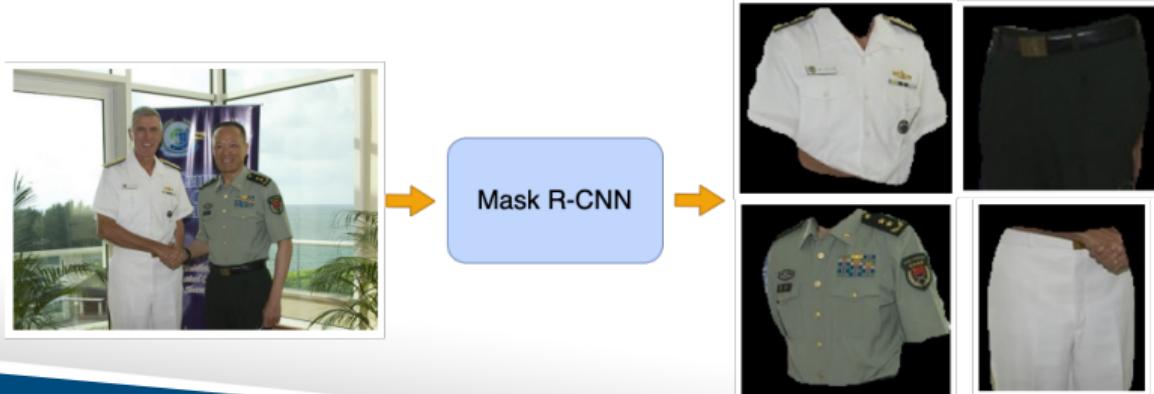
Step 1 - Mask R-CNN

- Based on Faster R-CNN
- **Three main parts**
- **Two parallel final branches** for classification + bounding box regression and segmentation



Step 2 - USED Preprocessing

- Use Mask R-CNN to make inference on all images of USED and create a new dataset
- For each detection create an image of the isolated (segmented) garment on a black background



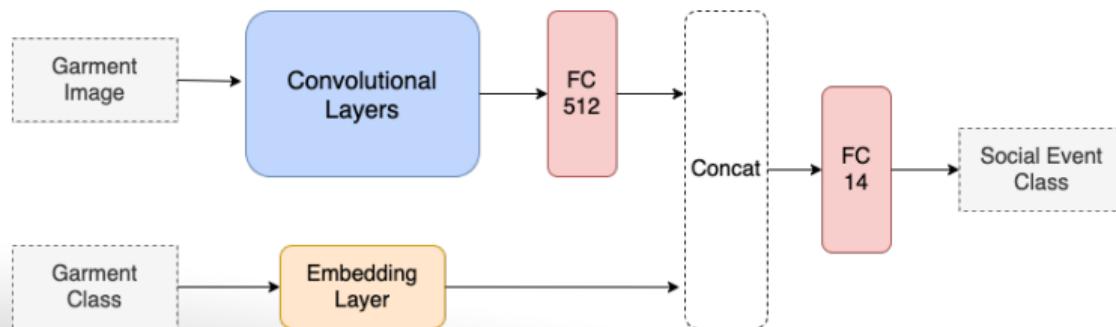
Step 3 - Events Classifier

- We used a **CNN-based classifier**
- We experimented with various architectures: VGG16, ResNet18 and **ResNet50** (the final choice)
- Final classifier part made of two fully-connected layers with 512 and 14 neurons respectively
- We started from a model pretrained on ImageNet, and we did a two-step training of **transfer learning + fine-tuning**



Step 3 - Events Classifier

- We built a second classifier using a **feature fusion approach**
 - The **garment class label** is sent as input to an **embedding layer**
 - The embedding layer output is concatenated to the image features from the convolutional part
- Additional info can improve the model performance



Experiments - Mask R-CNN Training

- We used **Detectron2** to build and train Mask R-CNN
- We replicated the training setup from the DeepFashion2 and Mask R-CNN papers
- Training took about 4 days on a Titan RTX GPU

Epochs	12
Batch Size	8
LR	$2 * 10^{-2}$ (*)
Optimizer	SGD (0.9 of momentum)
Weight Decay	10^{-5}

(*) Decreased by a factor of 10^{-1} at the 8th and the 11th epochs



Experiments - Mask R-CNN Results

- Model evaluated using the **COCO metrics**
- Our results were significantly lower than the DeepFashion2 authors
 - Authors trained on twice as many images
 - Training Mask R-CNN could improve the detection and segmentation performance

	Detection		Segmentation	
	Ours	Ge et al.	Ours	Ge et al.
AP	0.48	0.667	0.466	0.674
AP50	0.63	0.814	0.625	0.834
AP75	0.565	0.773	0.551	0.793



Experiments - Events Classifier Training

Problems Faced

- Dataset heavily unbalanced
- Strong presence of overfitting after few epochs

Solution

- Weighted Random Sampler: oversampling of minority classes and undersampling of majority ones
- Data augmentation techniques
- Regularization



Experiments - Events Classifier Training

- Training the classifier keeping the pretrained ResNet50 part frozen
- Unfreeze of the last convolutional part of ResNet and fine-tuning
- Monitor the validation loss and save the best models

	Transfer Learning	Fine-Tuning
Epochs	32	32
Batch Size	64	64
LR	$5 * 10^{-4}$	$5 * 10^{-5}$
Optimizer	Adam	Adam
Weight Decay	$1 * 10^{-4}$	$1 * 10^{-3}$



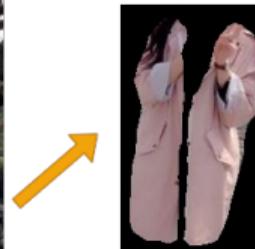
Experiments - Events Classifier Results

	Accuracy
ResNet50	0.494
ResNet50 + Emb.	0.499

ResNet50												ResNet50 With Embedding																	
concert	0.67	0.02	0.027	0.0044	0.013	0.013	0.006	0.0072	0.009	0.035	0.035	0.023	0.015	0.12	concert	0.71	0.011	0.016	0.0062	0.015	0.014	0.0032	0.0096	0.023	0.018	0.036	0.022	0.012	0.1
graduation	0.016	0.56	0.073	0.01	0.053	0.024	0.015	0.047	0.029	0.041	0.031	0.056	0.017	0.031	graduation	0.019	0.52	0.023	0.012	0.076	0.018	0.01	0.072	0.082	0.031	0.042	0.063	0.01	0.02
meeting	0.025	0.069	0.35	0.006	0.036	0.017	0.019	0.039	0.14	0.12	0.023	0.078	0.036	0.039	meeting	0.026	0.057	0.11	0.013	0.044	0.014	0.012	0.06	0.4	0.087	0.025	0.086	0.028	0.035
mountain-trip	0.016	0.014	0.01	0.55	0.079	0.066	0.057	0.0044	0.0038	0.029	0.0056	0.1	0.037	0.024	mountain-trip	0.012	0.016	0.0065	0.53	0.084	0.068	0.074	0.0059	0.016	0.019	0.016	0.11	0.034	0.013
picnic	0.014	0.05	0.059	0.022	0.39	0.083	0.022	0.019	0.011	0.049	0.033	0.12	0.087	0.044	picnic	0.016	0.034	0.028	0.053	0.43	0.077	0.017	0.028	0.04	0.036	0.034	0.11	0.062	0.029
sea-holiday	0.0084	0.02	0.018	0.028	0.09	0.52	0.047	0.016	0.011	0.041	0.044	0.057	0.057	0.042	sea-holiday	0.013	0.011	0.0091	0.057	0.12	0.49	0.036	0.026	0.014	0.029	0.046	0.069	0.064	0.018
ski-holiday	0.0082	0.011	0.016	0.027	0.014	0.03	0.78	0.0034	0.0056	0.027	0.009	0.035	0.023	0.013	ski-holiday	0.0071	0.014	0.0069	0.023	0.022	0.026	0.76	0.0034	0.02	0.016	0.013	0.059	0.019	0.0085
wedding	0.015	0.092	0.055	0.0059	0.032	0.021	0.0079	0.57	0.023	0.039	0.059	0.032	0.011	0.035	wedding	0.014	0.064	0.013	0.008	0.04	0.014	0.0044	0.63	0.067	0.025	0.061	0.028	0.0073	0.024
conference	0.025	0.075	0.36	0.0057	0.038	0.015	0.02	0.045	0.16	0.12	0.022	0.066	0.015	0.033	conference	0.022	0.062	0.11	0.012	0.047	0.012	0.011	0.066	0.43	0.085	0.025	0.08	0.0095	0.024
exhibition	0.032	0.056	0.15	0.012	0.049	0.05	0.026	0.031	0.058	0.27	0.066	0.1	0.033	0.061	exhibition	0.04	0.041	0.069	0.024	0.065	0.042	0.02	0.05	0.17	0.21	0.07	0.12	0.031	0.045
fashion	0.032	0.04	0.039	0.0079	0.033	0.033	0.018	0.041	0.0097	0.06	0.55	0.042	0.029	0.072	fashion	0.04	0.029	0.017	0.013	0.05	0.031	0.012	0.052	0.031	0.049	0.55	0.043	0.021	0.057
protest	0.02	0.047	0.065	0.021	0.077	0.047	0.035	0.014	0.013	0.065	0.022	0.48	0.051	0.042	protest	0.022	0.033	0.032	0.035	0.1	0.038	0.025	0.024	0.057	0.042	0.026	0.5	0.035	0.03
sport	0.0092	0.017	0.027	0.016	0.055	0.058	0.03	0.0052	0.0032	0.033	0.017	0.056	0.65	0.026	sport	0.013	0.013	0.023	0.022	0.08	0.04	0.026	0.0091	0.012	0.021	0.021	0.07	0.63	0.019
theater-dance	0.1	0.047	0.058	0.011	0.054	0.042	0.016	0.02	0.012	0.056	0.092	0.067	0.044	0.18	theater-dance	0.13	0.039	0.031	0.026	0.073	0.034	0.0091	0.043	0.045	0.039	0.1	0.069	0.033	0.33
	concert	graduation	meeting	mountain-trip	picnic	sea-holiday	ski-holiday	wedding	conference	exhibition	fashion	protest	sport	theater-dance		concert	graduation	meeting	mountain-trip	picnic	sea-holiday	ski-holiday	wedding	conference	exhibition	fashion	protest	sport	theater-dance



Application Example



Conclusions

- The proposed recommendation method gave good results and it is a good starting point for further research works
- Using images of single isolated garments could not give enough information to the network
- The embedding layer did not bring sensible improvements. This can be related to the garment type distribution among the dataset
- Future works may consider additional information related to the garment image (e.g. other clothes found in the same context)



Thank you for your attention



UNIVERSITÀ
DEGLI STUDI
FIRENZE