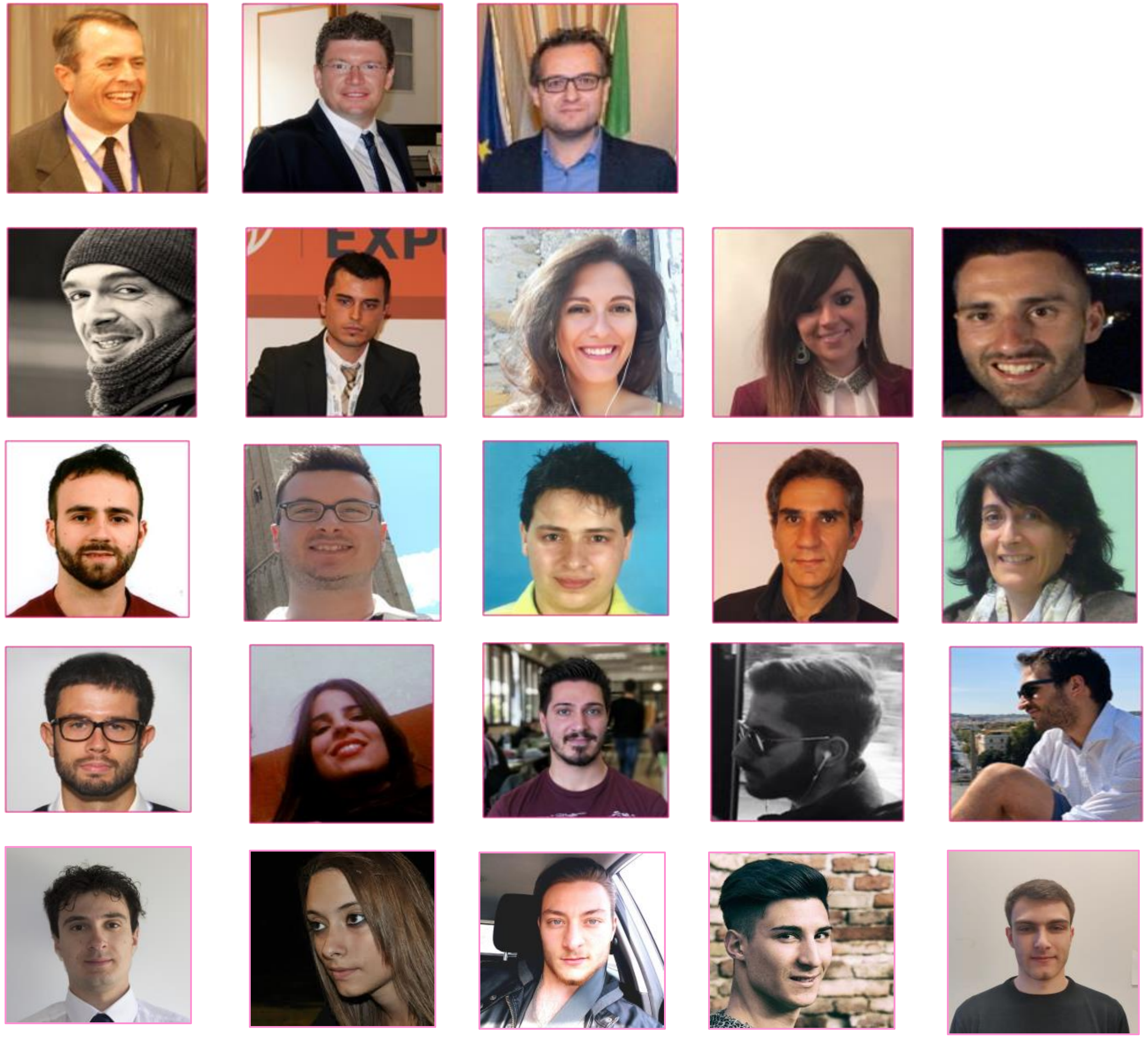
HOW TO DISCOVER TRENDS IN THE FASHION DOMAIN

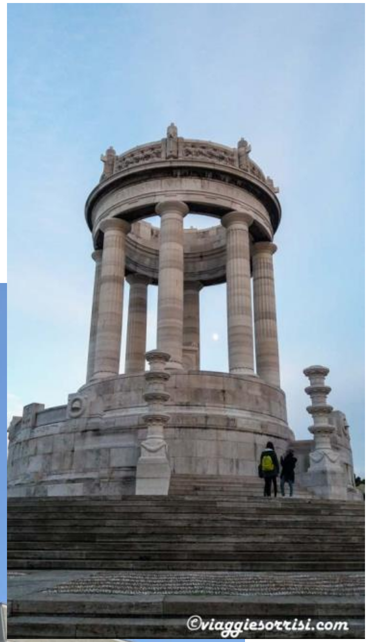
TALK LAYOUT

- VRAI @ UNIVPM
- Introduction
- Fashion Datasets
- Deep Learning for object detection
- Deep Learning for understanding and forecasting fashion trends
- UNIVPM Experiments
- Computational Time Comparison
- Use Cases – Real Examples
- Discussions and Open questions

VRAI @ UNIVPM



vision robotics
artificial
intelligence



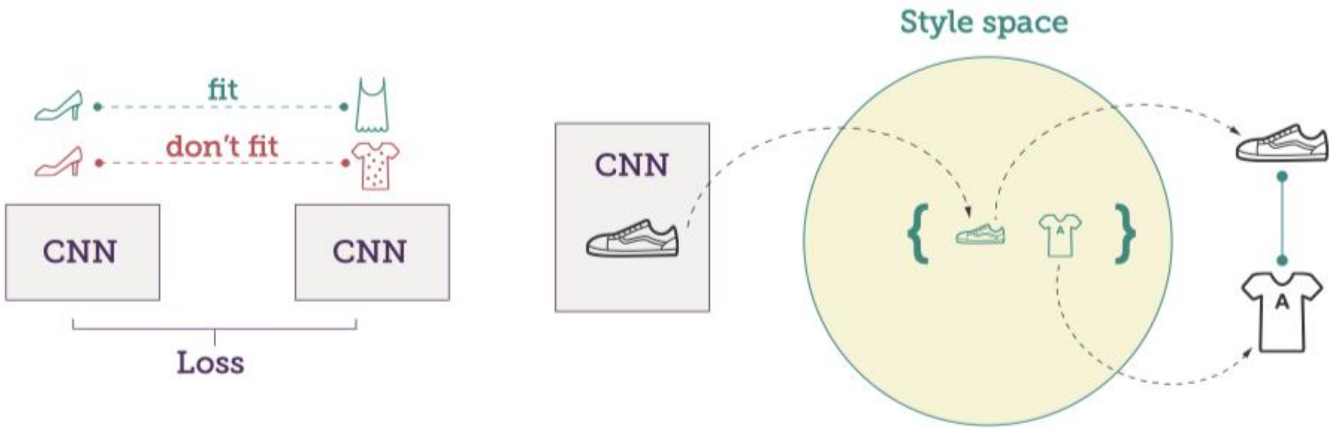
INTRODUCTION

Understanding fashion trends from images is certainly a goal that can bring numerous business benefits. It is an arduous task for designers, that try to get a **summary of trends.**

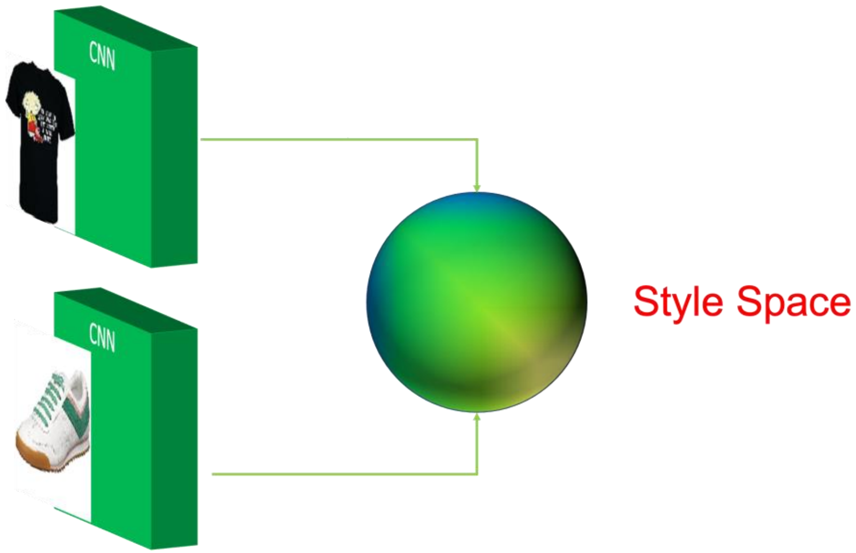
However, analyzing **large image dataset** is difficult using features-based techniques.

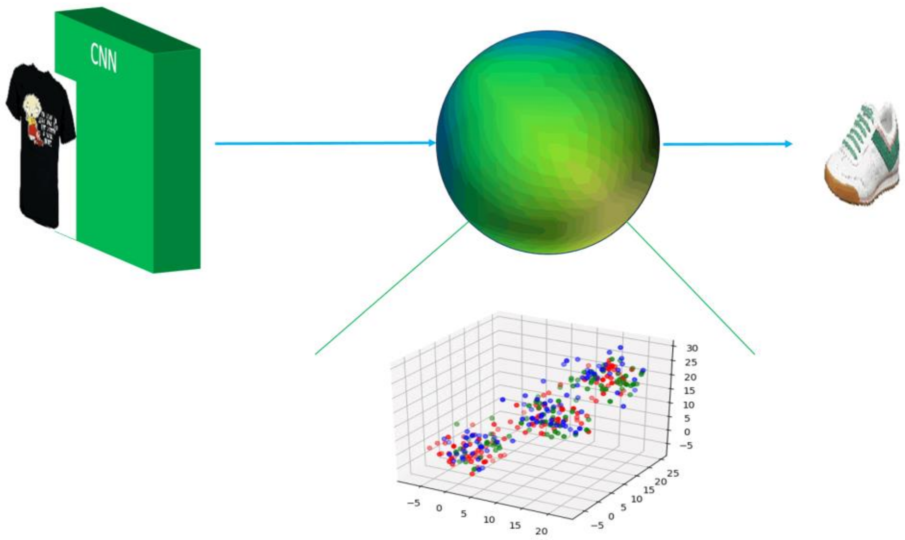
In recent years, deep learning speeds up this process and algorithms that **detect clothes** from images are specifically developed.

INTRODUCTION



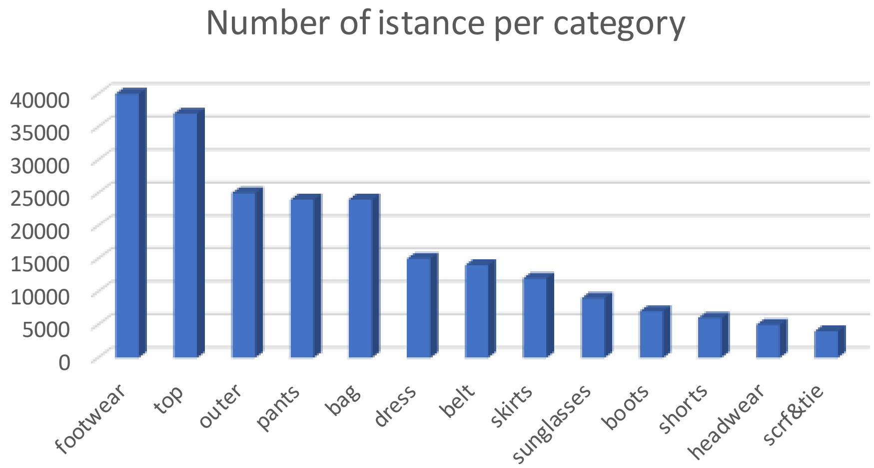
To answer questions like ‘What outfit goes well with this pair of shoes?’, one has to go beyond learning visual similarity and learn a visual notion of compatibility across categories. Learning frameworks are developed to help answering these types of questions.

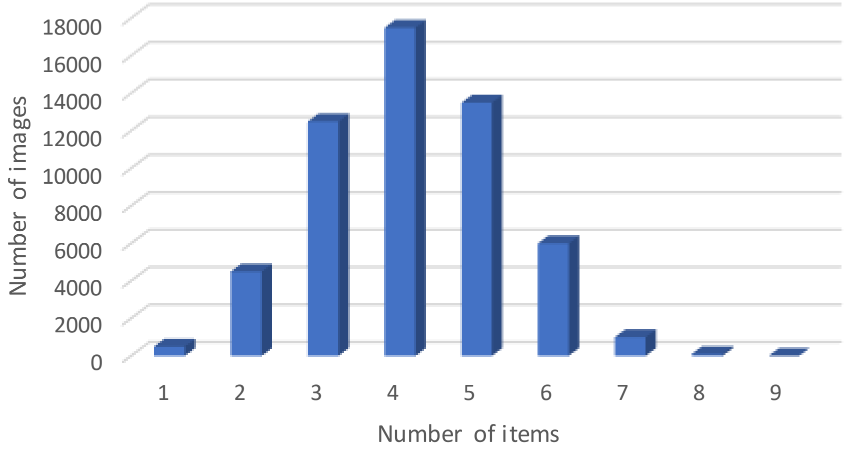




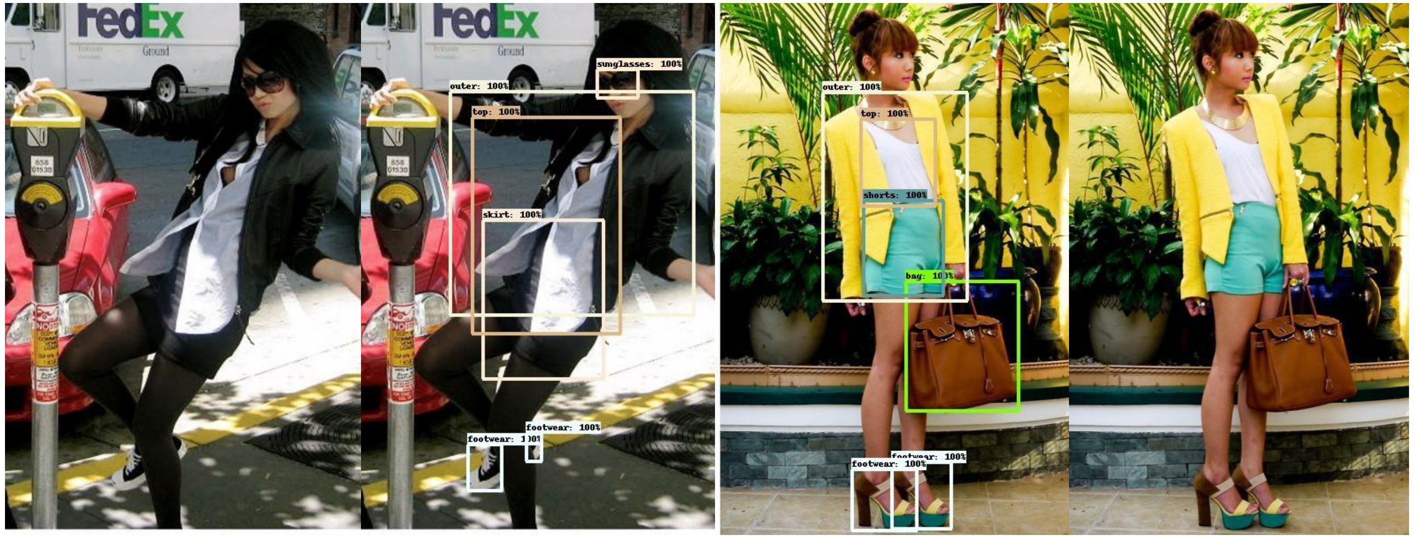
Veit, A., Kovacs, B., Bell, S., McAuley, J., Bala, K., & Belongie, S. (2015). Learning visual clothing style with heterogeneous dyadic co-occurrences. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4642-4650).

MODANET DATASET





The **Modanet** dataset contains 55176 images fully annotated which are divided into 52377 images for training set and 2799 images for the validation set.



Zheng, S., Yang, F., Kiapour, M. H., & Piramuthu, R. (2018, October). Modanet: A large-scale street fashion dataset with polygon annotations. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 1670-1678).

FASHION-MNIST DATASET

Fashion-MNIST is a dataset of Zalando's article images consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 grayscale image, associated with a label from 10 classes. Fashion-MNIST serves as a direct drop-in replacement for the original MNIST dataset for benchmarking machine learning algorithms. It shares the same image size and structure of training and testing splits.

Bhatnagar, S., Ghosal, D., & Kolekar, M. H. (2017, December). Classification of fashion article images using convolutional neural networks. In 2017 Fourth International Conference on Image Information Processing (ICIIP) (pp. 1-6). IEEE.

DEEP FASHION DATASET

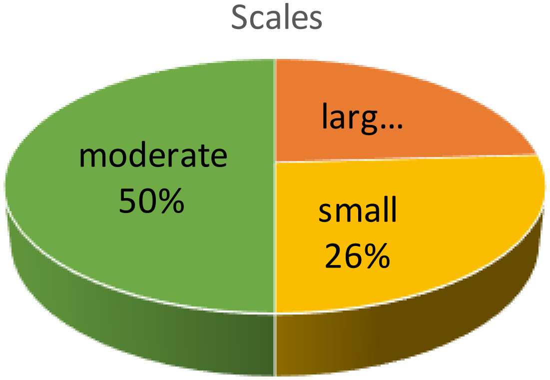
Deep-Fashion Dataset is a large-scale clothes database. **Four benchmarks** are developed using the DeepFashion database, including **Attribute Prediction**, **Consumer-to-shop Clothes Retrieval**, **In-shop Clothes Retrieval**, and **Landmark Detection**. The data and annotations of these benchmarks can be also employed as the training and test sets for the following computer vision tasks, such as Clothes Detection, Clothes Recognition, and Image Retrieval.

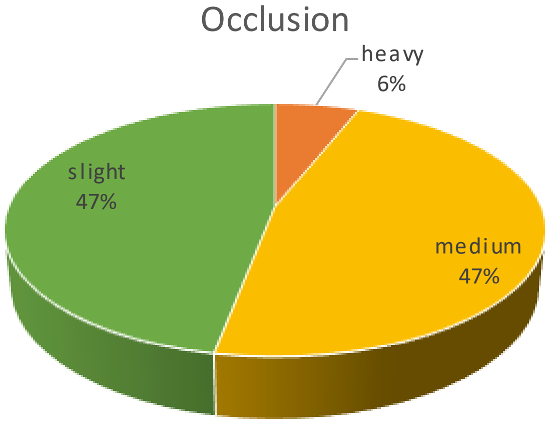
- 800000 imgs
  - 50 categories
  - 1000
    - Attributes
    - Bounding box
    - Landmark
  - 300000
    - Cross-pose
    - Cross-domain

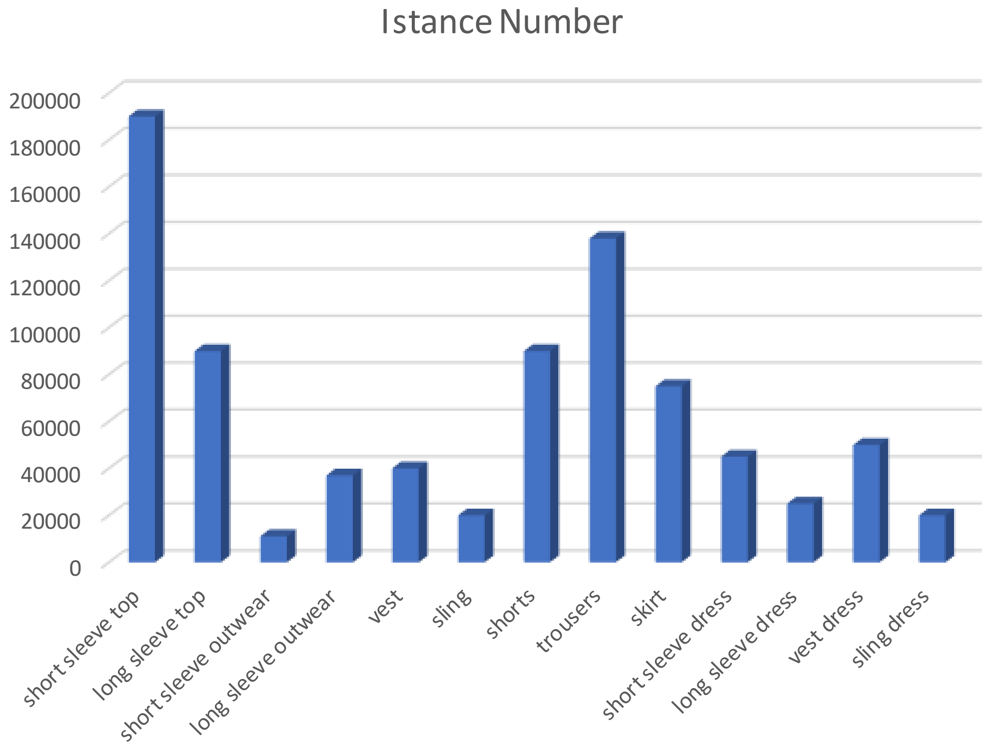
Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1096-1104).

DEEP-FASHION 2 DATASET

Deep Fashion 2 is a comprehensive fashion dataset. It contains 491K diverse images of 13 popular clothing categories from both commercial shopping stores and consumers. It totally comprises 801K clothing items, where each item in an image labelled with scale, occlusion, zoom-in, viewpoint, category, style, bounding box, dense landmarks and per-pixel mask. There are also 873K Commercial-Consumer clothes pairs.

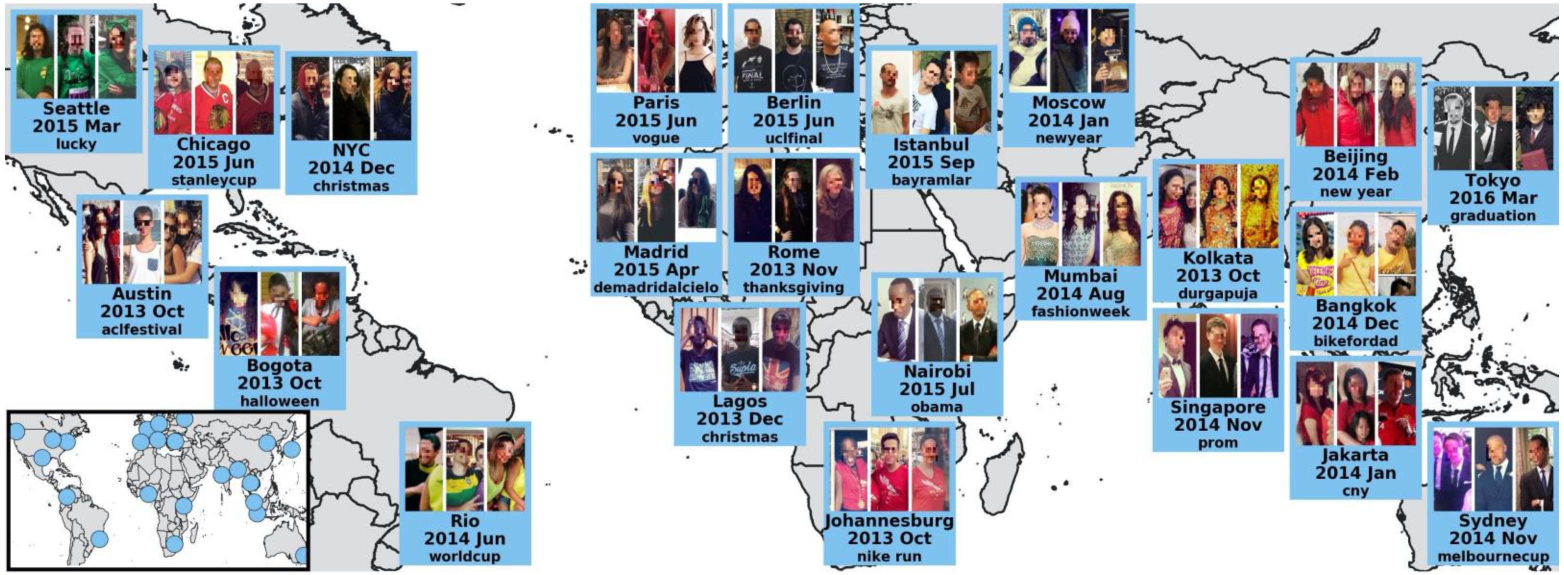






Ge, Y., Zhang, R., Wang, X., Tang, X., & Luo, P. (2019). Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5337-5345)

GeoStyle DATASET

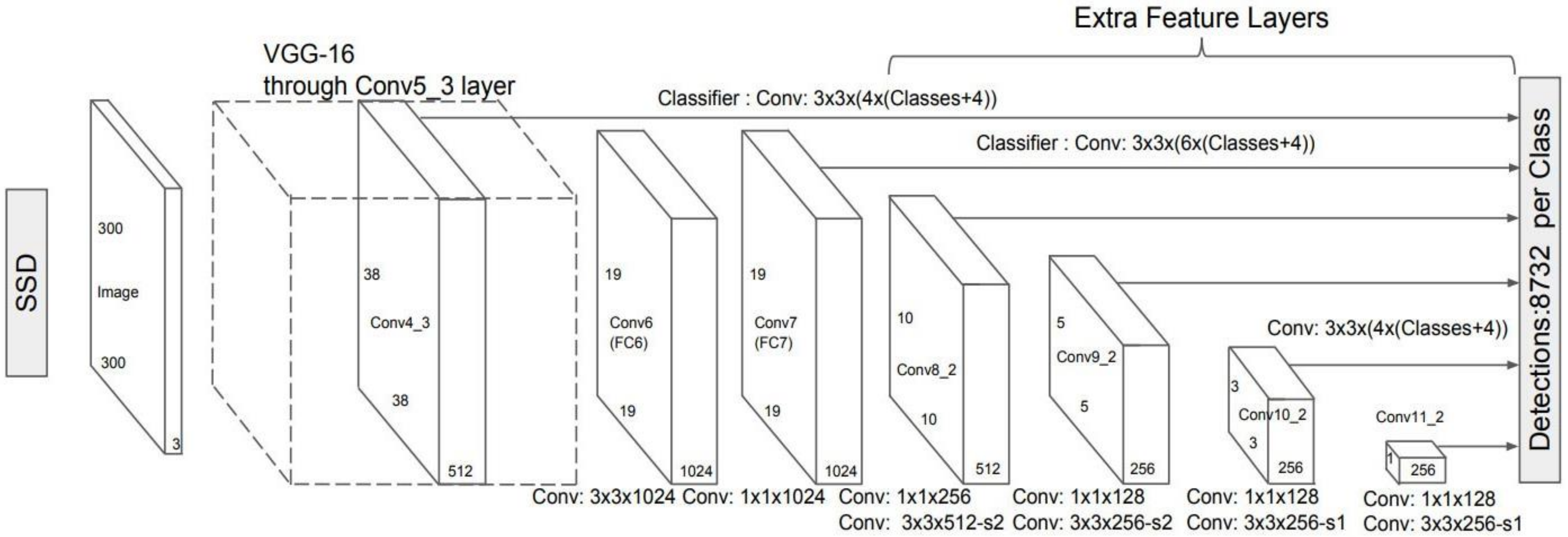


Based on Instagram and Flickr photos showing people from 44 major cities from around the world. The dataset has 7.7 million images that span a time period from July 2013 until May 2016. The dataset is used for research purposes only.

Mall, U., Matzen, K., Hariharan, B., Snavely, N., & Bala, K. (2019). Geostyle: Discovering fashion trends and events. In Proceedings of the IEEE International Conference on Computer Vision (pp. 411-420).

SINGLE SHOT MULTIBOX DETECTOR

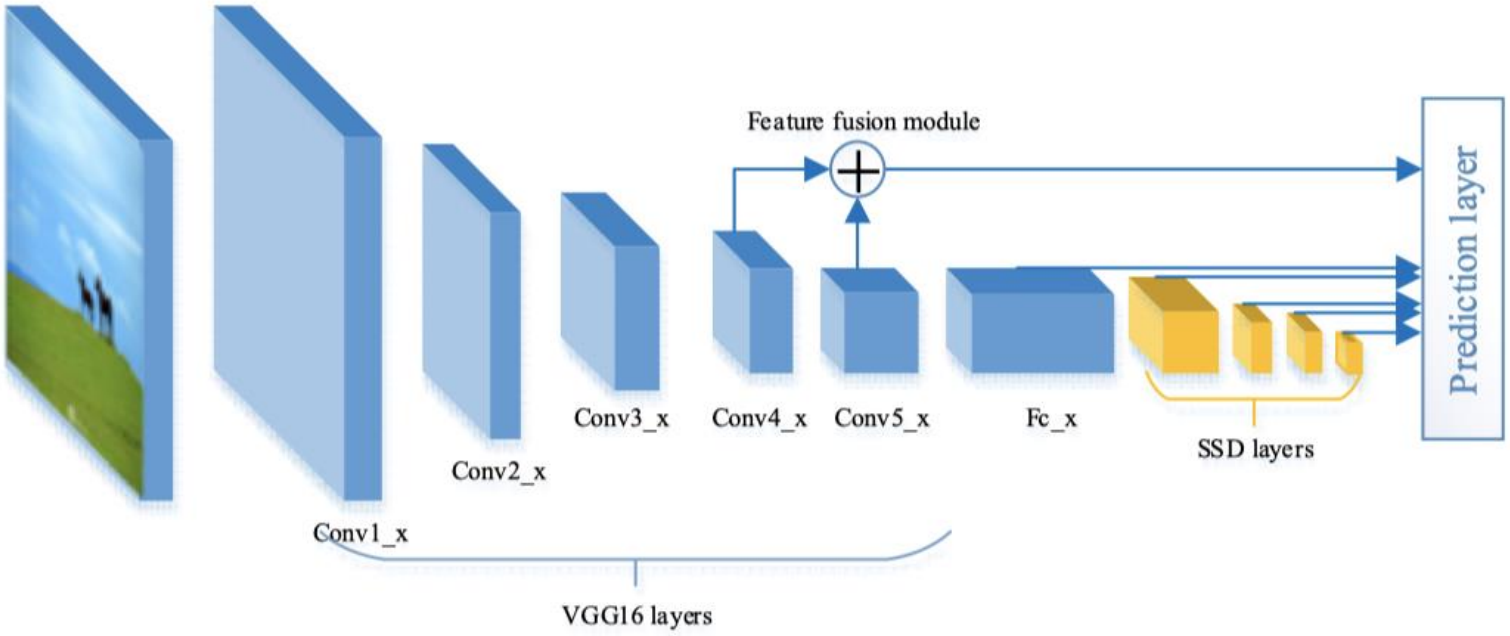
A single deep neural network it is used for detecting objects in images. This approach, named SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location.



Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.

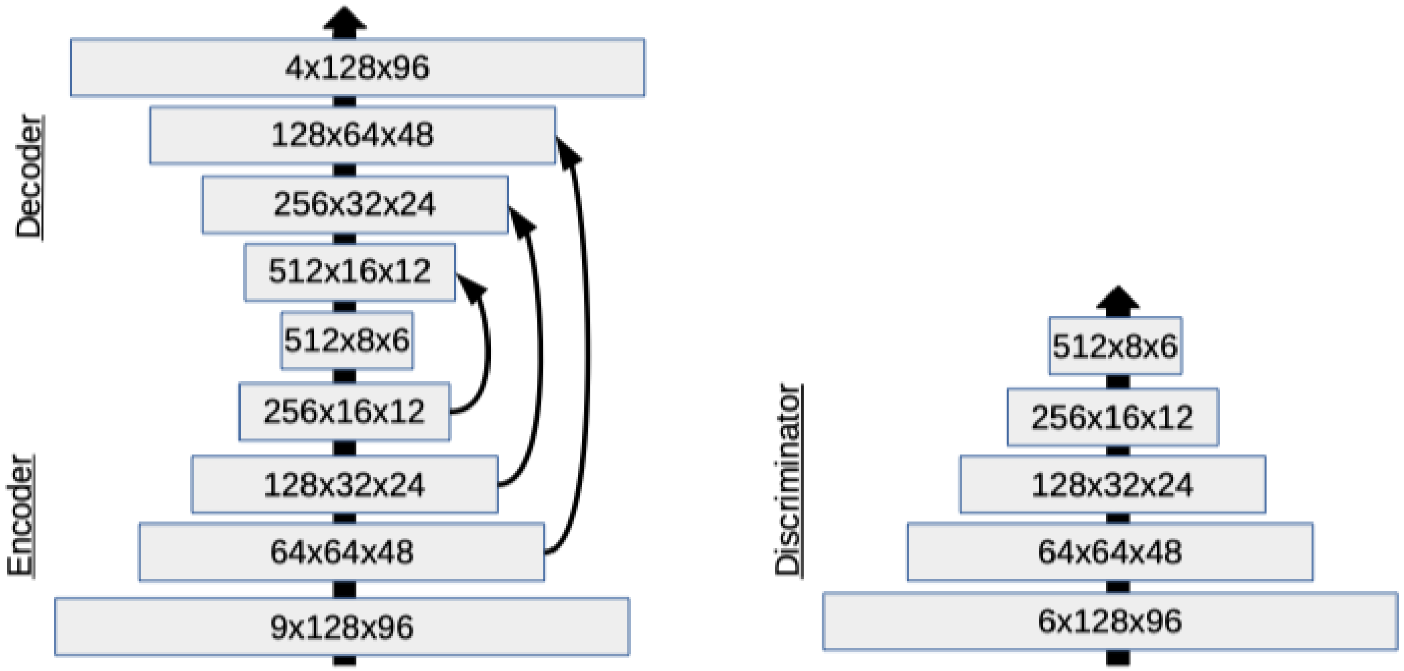
FEATURE-FUSED SSD

A multi-level feature fusion method for introducing contextual information in SSD, is developed,in order to improve the accuracy for small objects. Small objects detection is a challenging task in computer vision due to its limited resolution and information. The majority of existing methods sacrifice speed for improvement in accuracy. Feature-fused SSD detects small objects at a fast speed, using the best object detector Single Shot Multibox Detector (SSD) with respect to accuracy-vs-speed trade-off as base architecture.



Cao, G., Xie, X., Yang, W., Liao, Q., Shi, G., & Wu, J. (2018, April). Feature-fused SSD: fast detection for small objects. In Ninth International Conference on Graphic and Image Processing (ICGIP 2017) (Vol. 10615, p. 106151E). International Society for Optics and Photonics.

CONDITIONAL ANALOGY GENERATIVE ADVERSARIAL NETWORK

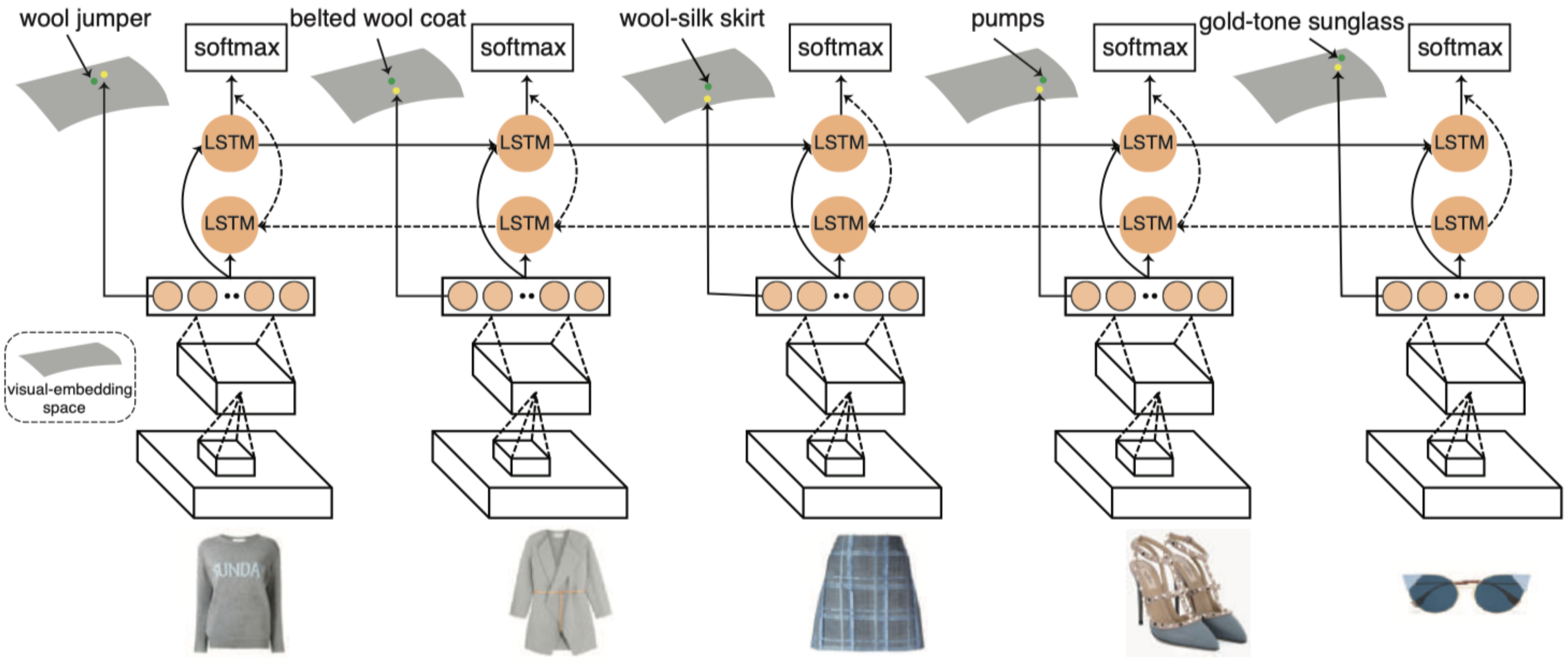


Method to solve image analogy problems: it allows to learn the relation between paired images present in training data, and then generalize and generate images that correspond to the relation but were never seen in the training set. Conditional Analogy Generative Adversarial Network (CAGAN) is based on adversarial training and employs deep convolutional neural networks. An especially interesting application of that technique is automatic swapping of clothing on fashion model photos.

Jetchev, Nikolay, and Urs Bergmann. "The conditional analogy gan: Swapping fashion articles on people images." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017.

LEARNING FASHION COMPATIBILITY WITH BI-LSTM

Two types of fashion recommendation are studied: (i) suggesting an item that matches existing components in a set to form a stylish outfit (a collection of fashion items), and (ii) generating an outfit with multimodal (images/text) specifications from a user. To this end, a jointly learn a visual-semantic embedding and the compatibility relationships among fashion items in an end-to-end fashion is proposed.



Han, X., Wu, Z., Jiang, Y. G., & Davis, L. S. (2017, October). Learning fashion compatibility with bidirectional lstms. In Proceedings of the 25th ACM international conference on Multimedia (pp. 1078-1086).

LEARNING FASHION COMPATIBILITY WITH BI-LSTM

Outfit
Valuation

Outfit
Completion

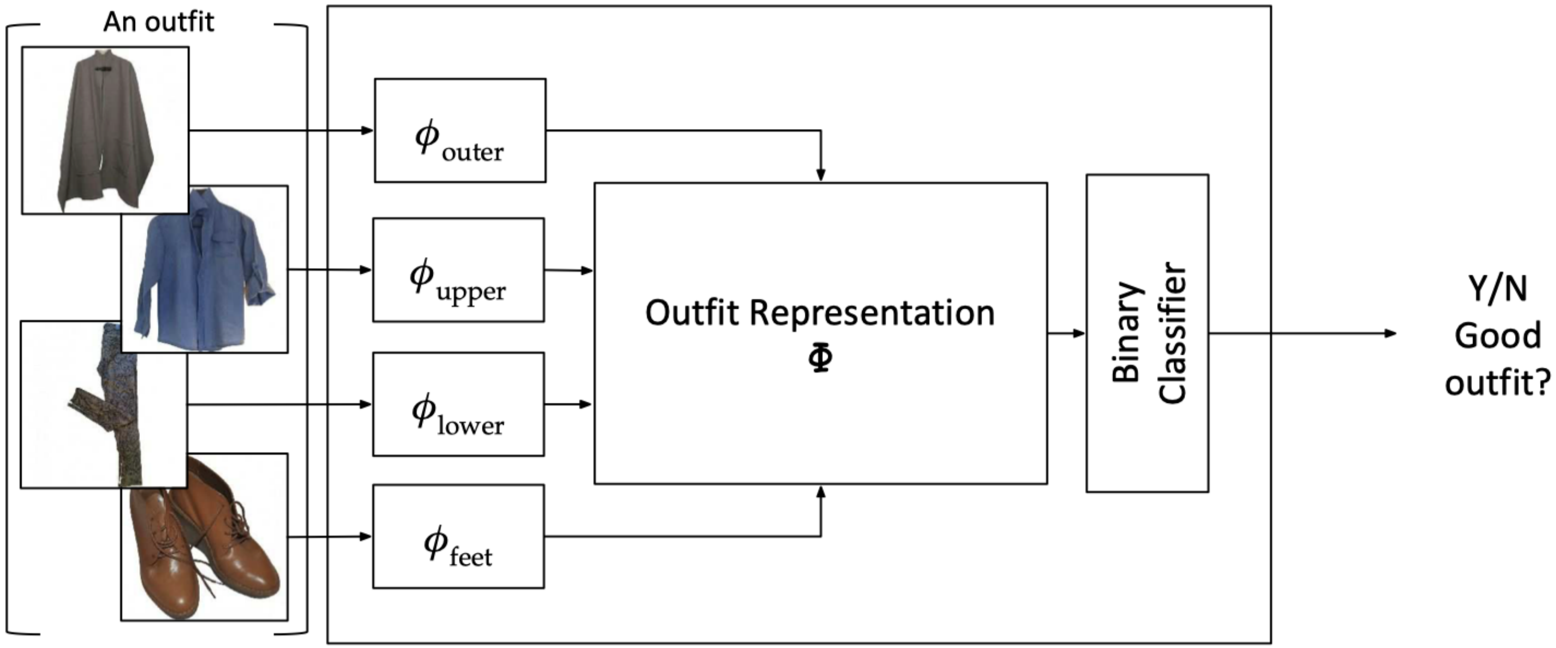
Outfit
Generation

Denim

Han, X., Wu, Z., Jiang, Y. G., & Davis, L. S. (2017, October). Learning fashion compatibility with bidirectional lstms. In Proceedings of the 25th ACM international conference on Multimedia (pp. 1078-1086).

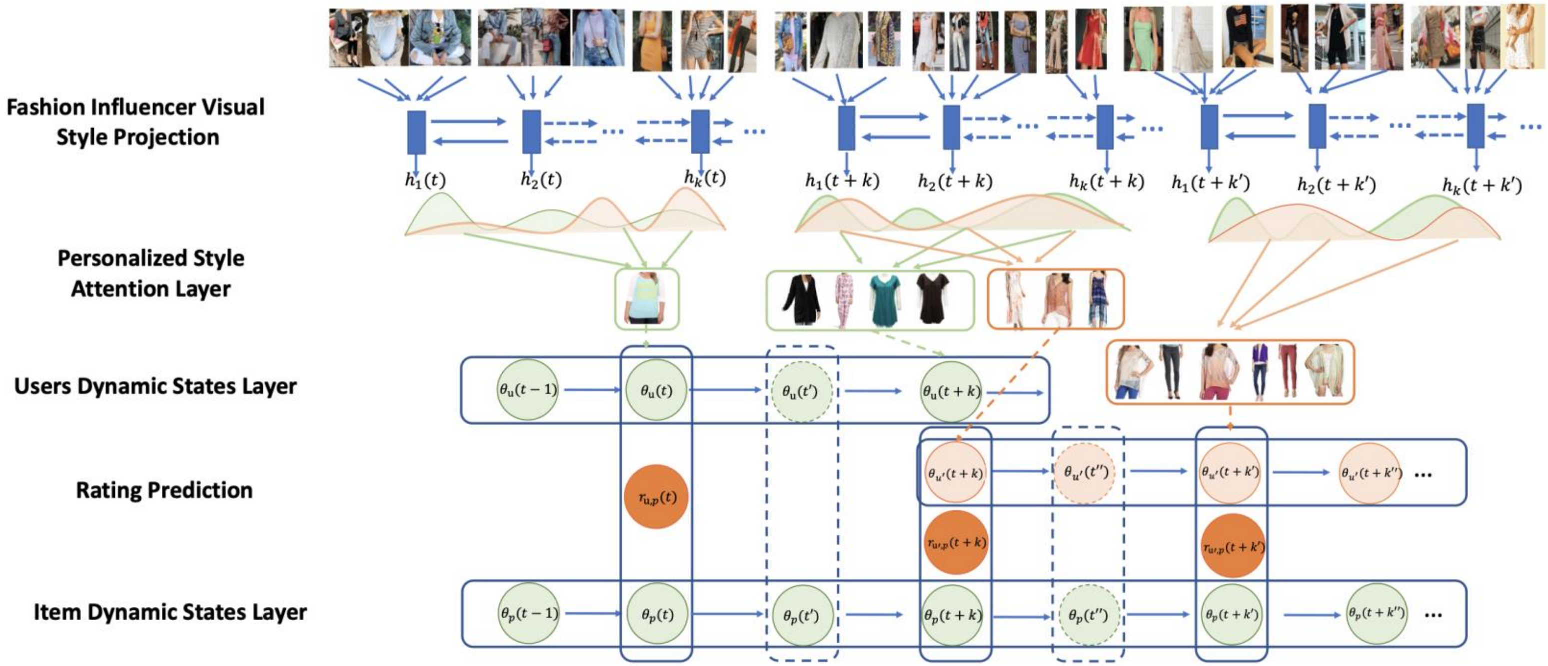
RECOMMENDING OUTFIT FROM PERSONAL CLOSET

The outfit grading problem is considered for outfit recommendation, where it is assumed that users have a closet of items and the aim is producing a score for an arbitrary combination of items in the closet. The challenge in outfit grading is that the input to the system is a bag of item pictures that are unordered and vary in size. For this reason, a deep neural network-based system is built that can take variablelength items and predict a score.



Tangseng, P., Yamaguchi, K., & Okatani, T. (2017). Recommending outfits from personal closet. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 2275-2279).

RECURRENT FASHION RECOMMENDATION

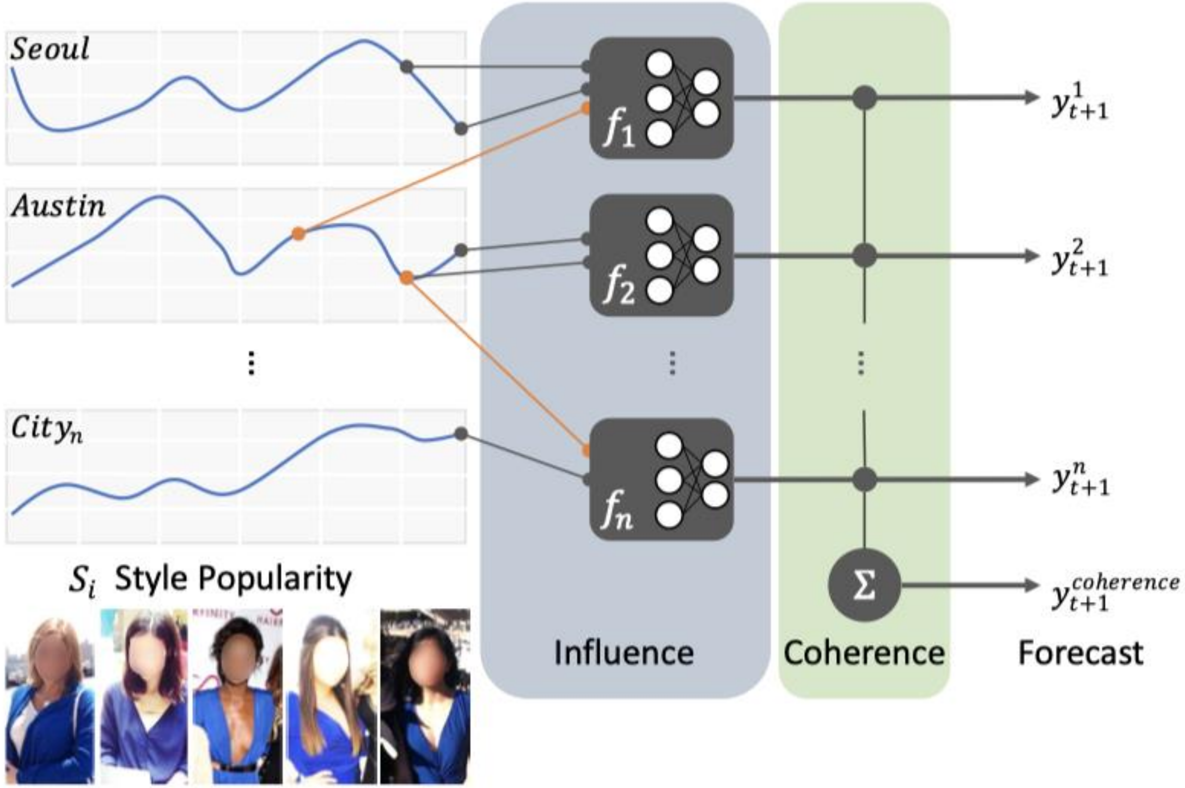


Fashion-focused key opinion bloggers on Instagram, Facebook, and other social media platforms are fast becoming critical influencers. They can inspire consumer clothing purchases by linking high fashion visual evolution with daily street style. Based on this assumption, it it build the first visual influence-aware fashion recommender (FIRN) with leveraging fashion bloggers and their dynamic visual posts. Specifically, it is extracted the dynamic fashion features highlighted by these bloggers via a BiLSTM that integrates a large corpus of visual posts and community influence.

Zhang, Y., & Caverlee, J. (2019, November). Instagrammers, fashionistas, and me: Recurrent fashion recommendation with implicit visual influence. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 1583-1592).

DISCOVERING FASHION STYLE

The evolution of clothing styles and their migration across the world is intriguing, yet difficult to describe quantitatively. An approach to discover and quantify fashion influences from everyday images of people wearing clothes is proposed. This detects which cities influence which other cities in terms of propagating their styles.



Al-Halah, Z., & Grauman, K. (2020). From Paris to Berlin: Discovering Fashion Style Influences Around the World. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10136-10145).

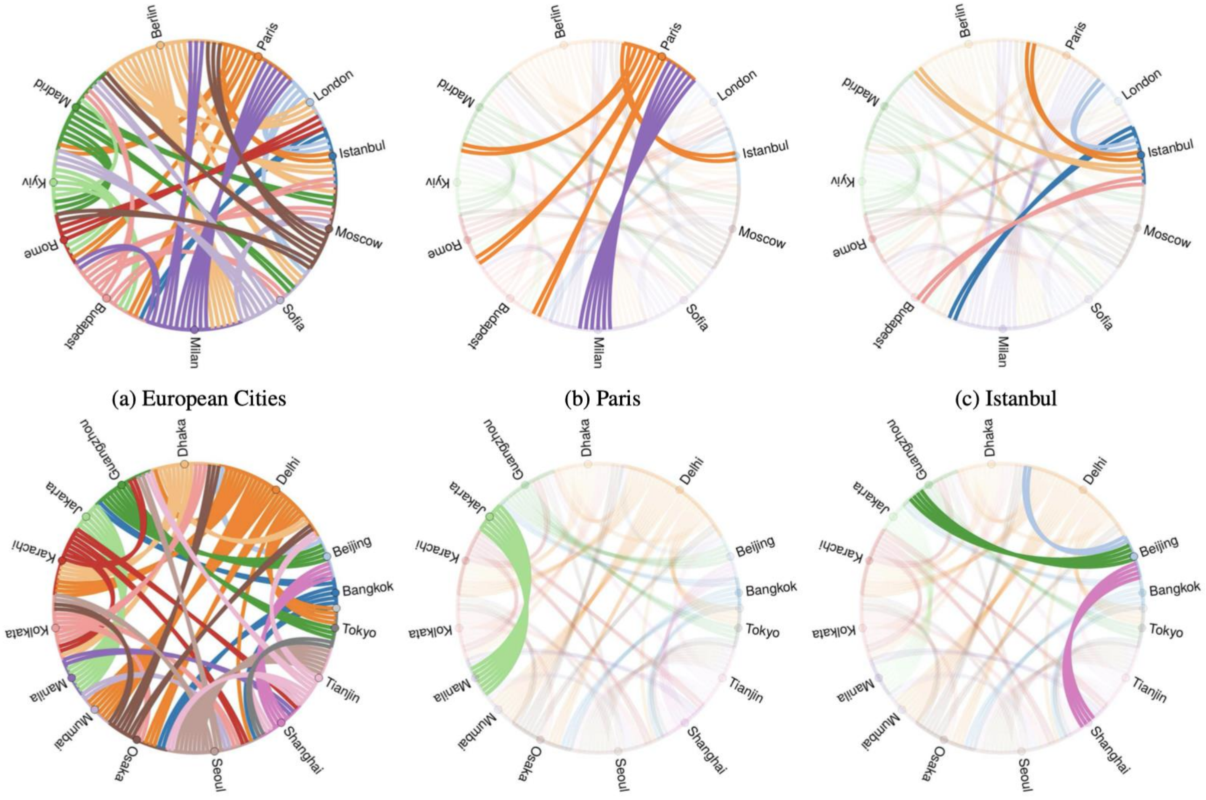
DISCOVERING FASHION STYLE

Trend

Images					
City	Bangkok	Seattle	Bangkok	Chicago	Rio
Attribute	Yellow color	No sleeves	T-shirt	Red color	Yellow color
Keywords	dad, father	halloween, freaknight	songkran, festival	cup, stanleycup	worldcup, brasil

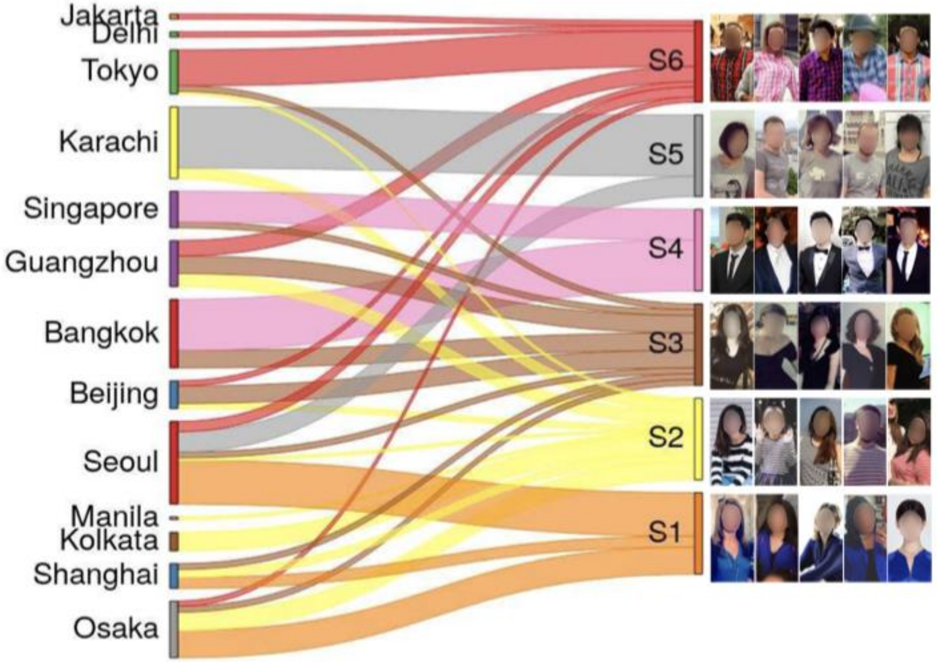
DISCOVERING FASHION STYLE

Influence

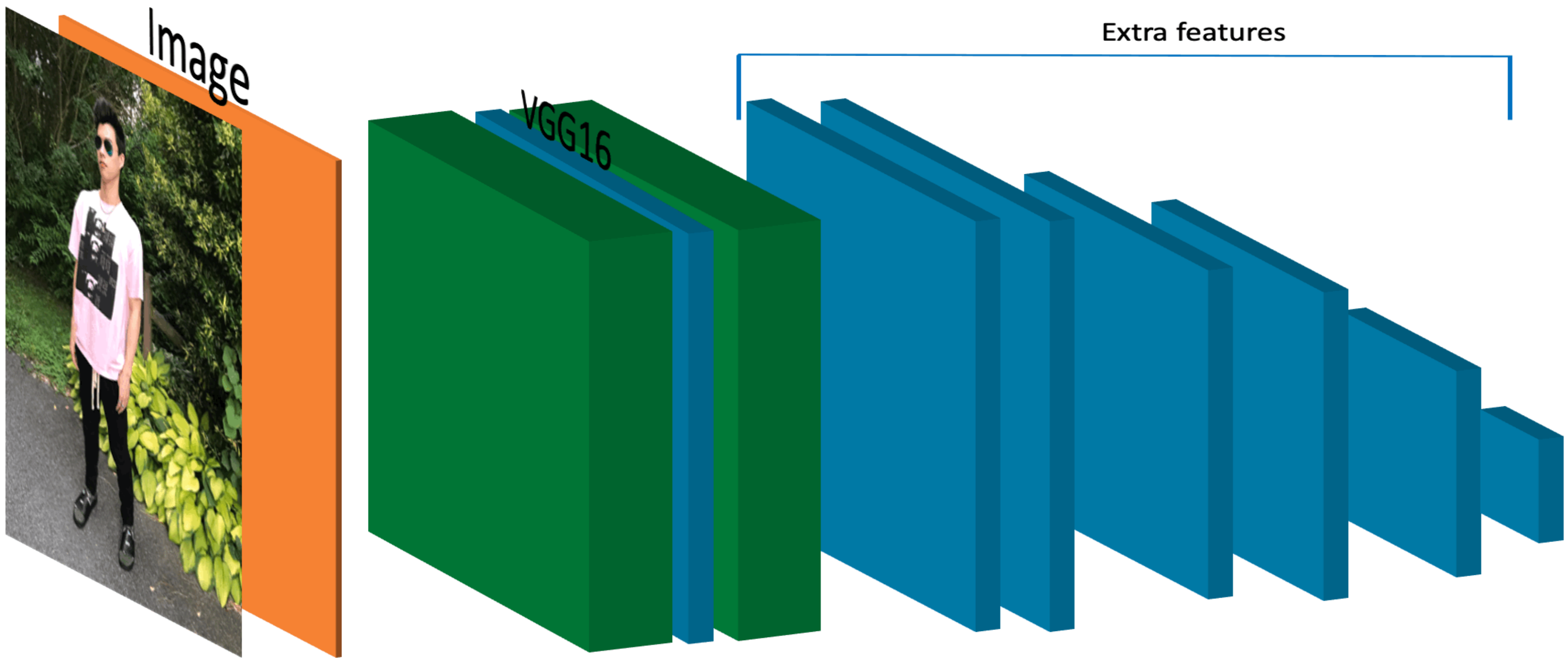


(a) European Cities (b) Paris (c) Istanbul

(d) Asian Cities (e) Jakarta (f) Beijing

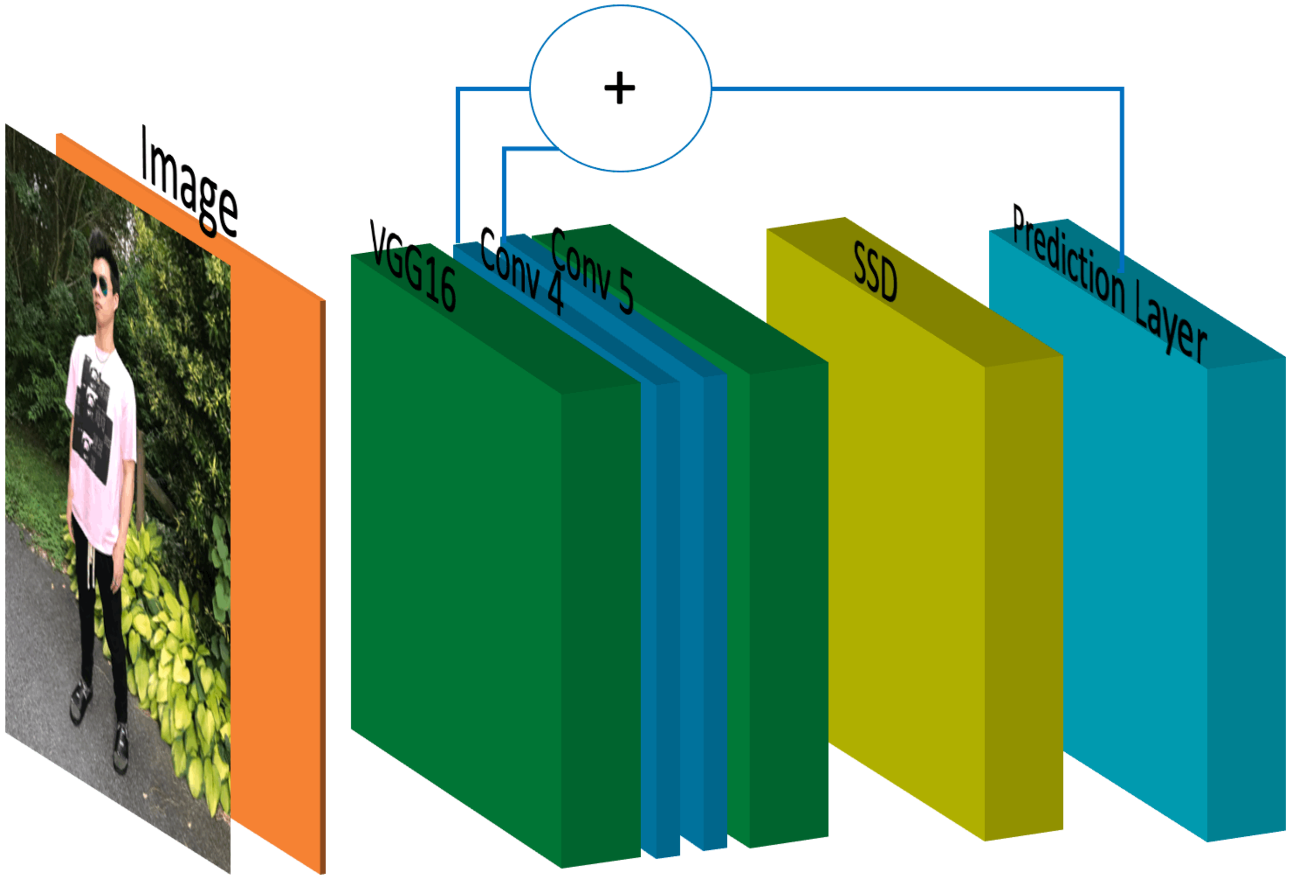


UNIVPM EXPERIMENTS



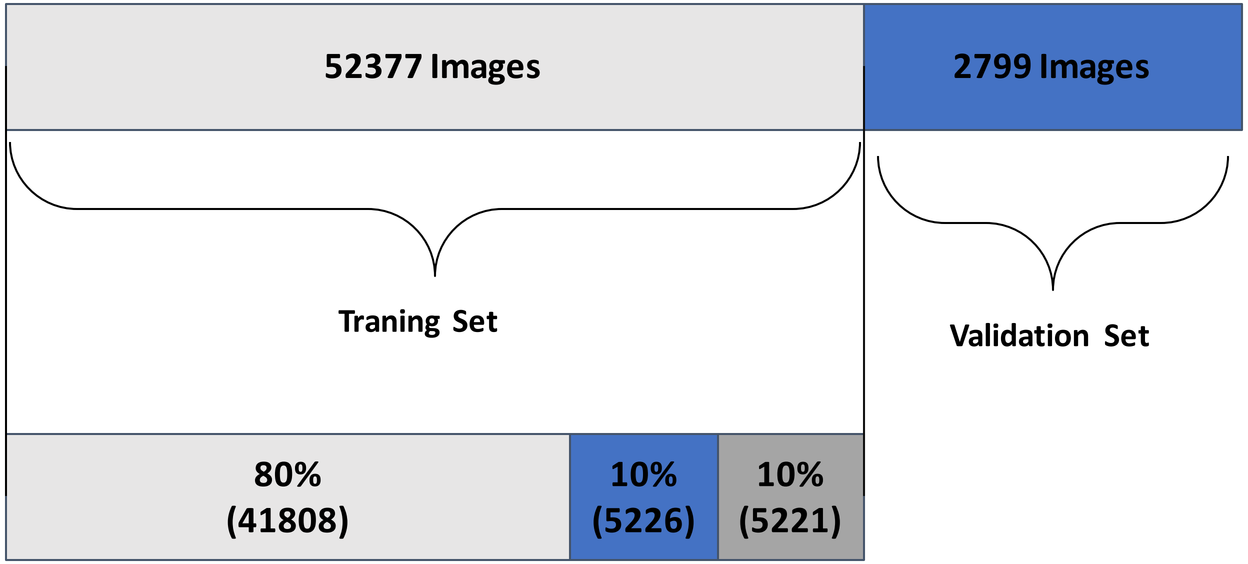
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.

UNIVPM EXPERIMENTS

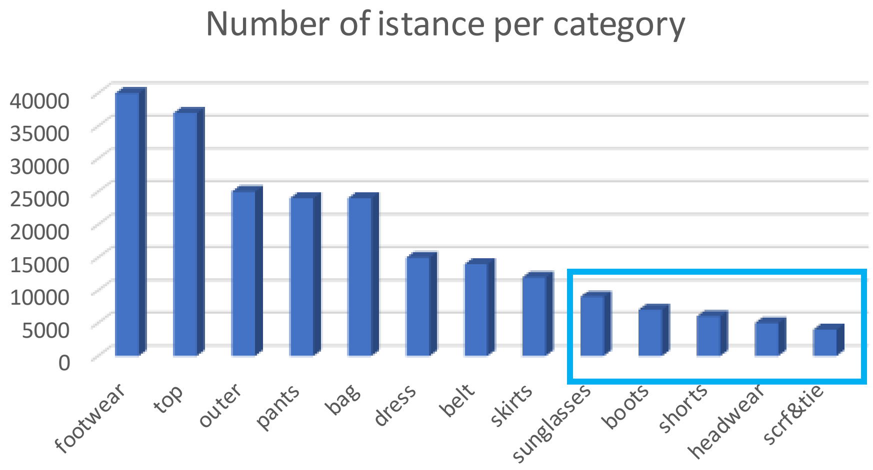


Cao, G., Xie, X., Yang, W., Liao, Q., Shi, G., & Wu, J. (2018, April). Feature-fused SSD: fast detection for small objects. In Ninth International Conference on Graphic and Image Processing (ICGIP 2017) (Vol. 10615, p. 106151E). International Society for Optics and Photonics.

UNIVPM EXPERIMENTS



The dataset choose for the experiment is Modanet. In particular, the training set is splitted in Training, Validation and Test Set because the Validation set of this dataset is not annotated.



Due to the imbalance presence the dataset is splitted manually.

UNIVPM EXPERIMENTS

Model configuration parameters

```
img_height = 300 # Height of the model input images
img_width = 300 # Width of the model input images
img_channels = 3 # Number of color channels of the model i
mean_color = [123, 117, 104] # The per-channel mean of the
swap_channels = [2, 1, 0] # The color channel order in the
n_classes = 13
scales_coco = [0.07, 0.15, 0.33, 0.51, 0.69, 0.87, 1.05] #
scales = scales_coco
aspect_ratios = [[1.0, 2.0, 0.5],
                 [1.0, 2.0, 0.5, 3.0, 1.0/3.0],
                 [1.0, 2.0, 0.5, 3.0, 1.0/3.0],
                 [1.0, 2.0, 0.5, 3.0, 1.0/3.0],
                 [1.0, 2.0, 0.5],
                 [1.0, 2.0, 0.5]] # The anchor box aspect
```

To use the same aspect ratio of Instragram photos

The training parameters

Batch size is 64, for 120 epoch with 256 step per epoch.

The learning rate is scheduled by epoch ranges:

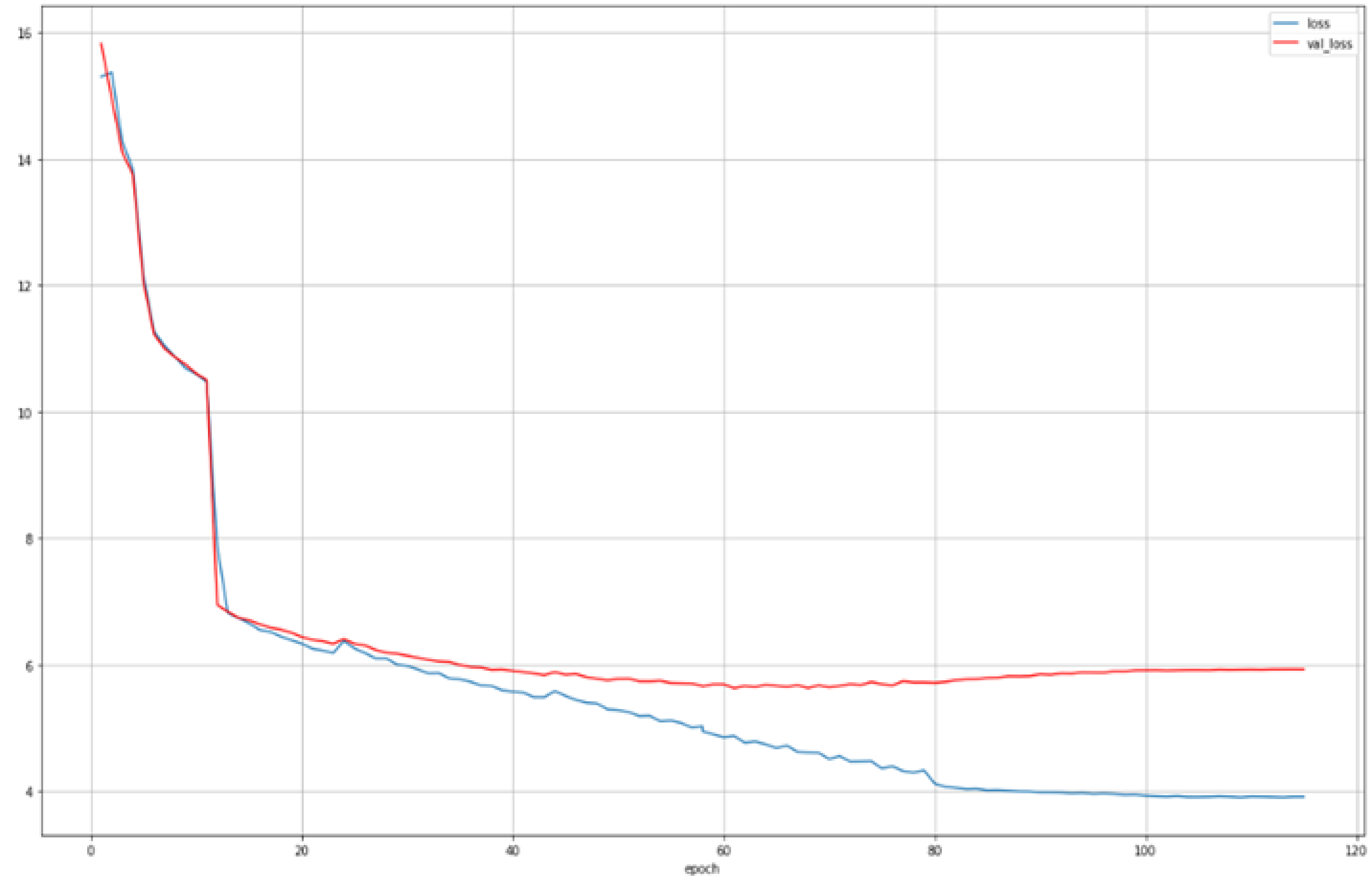
- for the epoch < 80 —> lr=0.001
- for the epoch < 100 —> lr = 0.0001
- for the epoch > 100 —> lr = 0.00001

UNIVPM EXPERIMENTS

Training Evaluation for SSD

- Batch size 64
- 120 epoch
- 256 iteration per epoch
- Optimizer SGD

In the training phase the best epoch was 62.

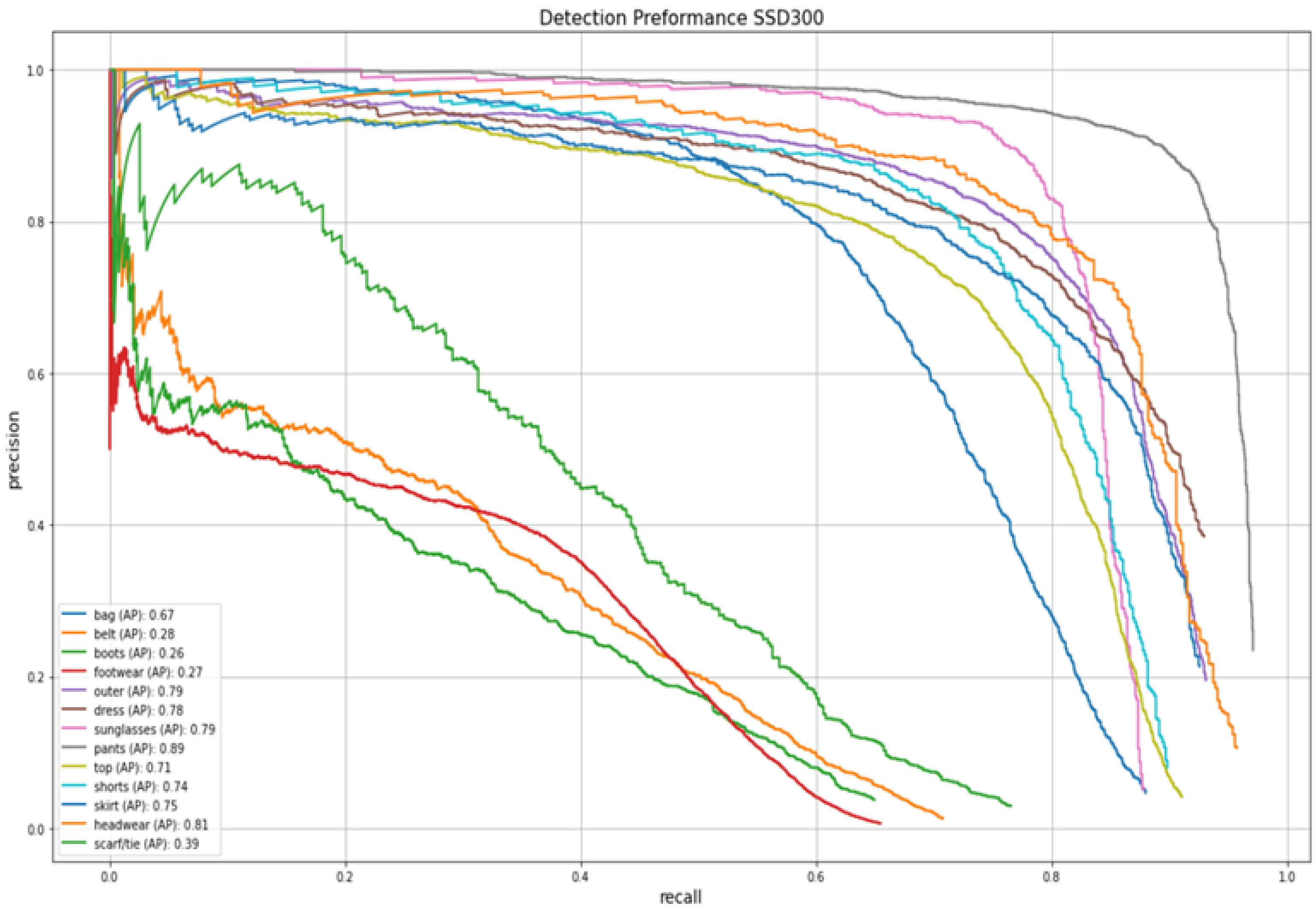


UNIVPM EXPERIMENTS

Test evaluation for SSD

The evaluation confirms that for partially occluded and small clothes the neural network has low performance.

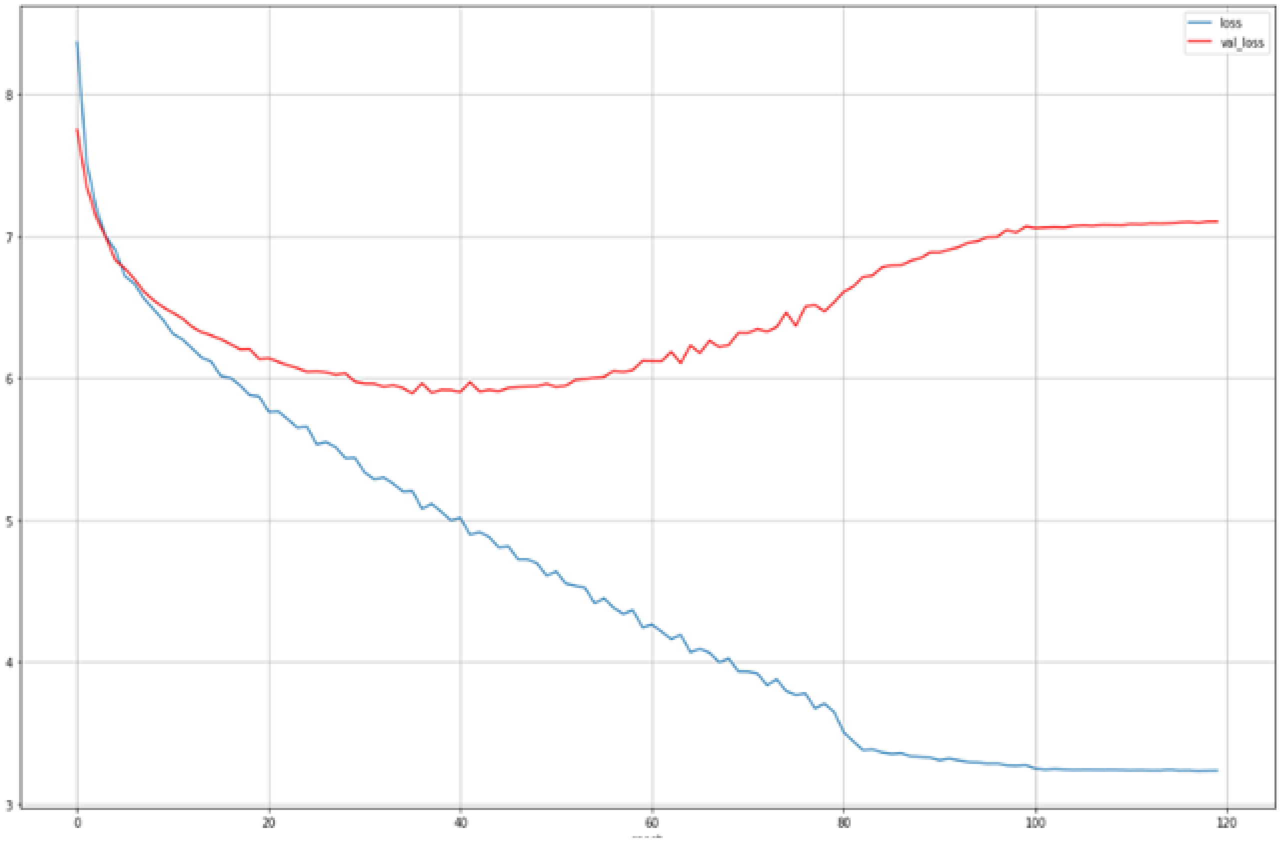
bag	AP	0.674
belt	AP	0.285
boots	AP	0.26
footwear	AP	0.27
outer	AP	0.785
dress	AP	0.784
sunglasses	AP	0.79
pants	AP	0.885
top	AP	0.705
shorts	AP	0.74
skirt	AP	0.754
headwear	AP	0.809
scarf/tie	AP	0.387
All categories	mAP	0.625



UNIVPM EXPERIMENTS

Training valuation for FFSSD

- Batch size 64
- 120 epoch
- 256 iteration per epoch
- Optimizer SGD

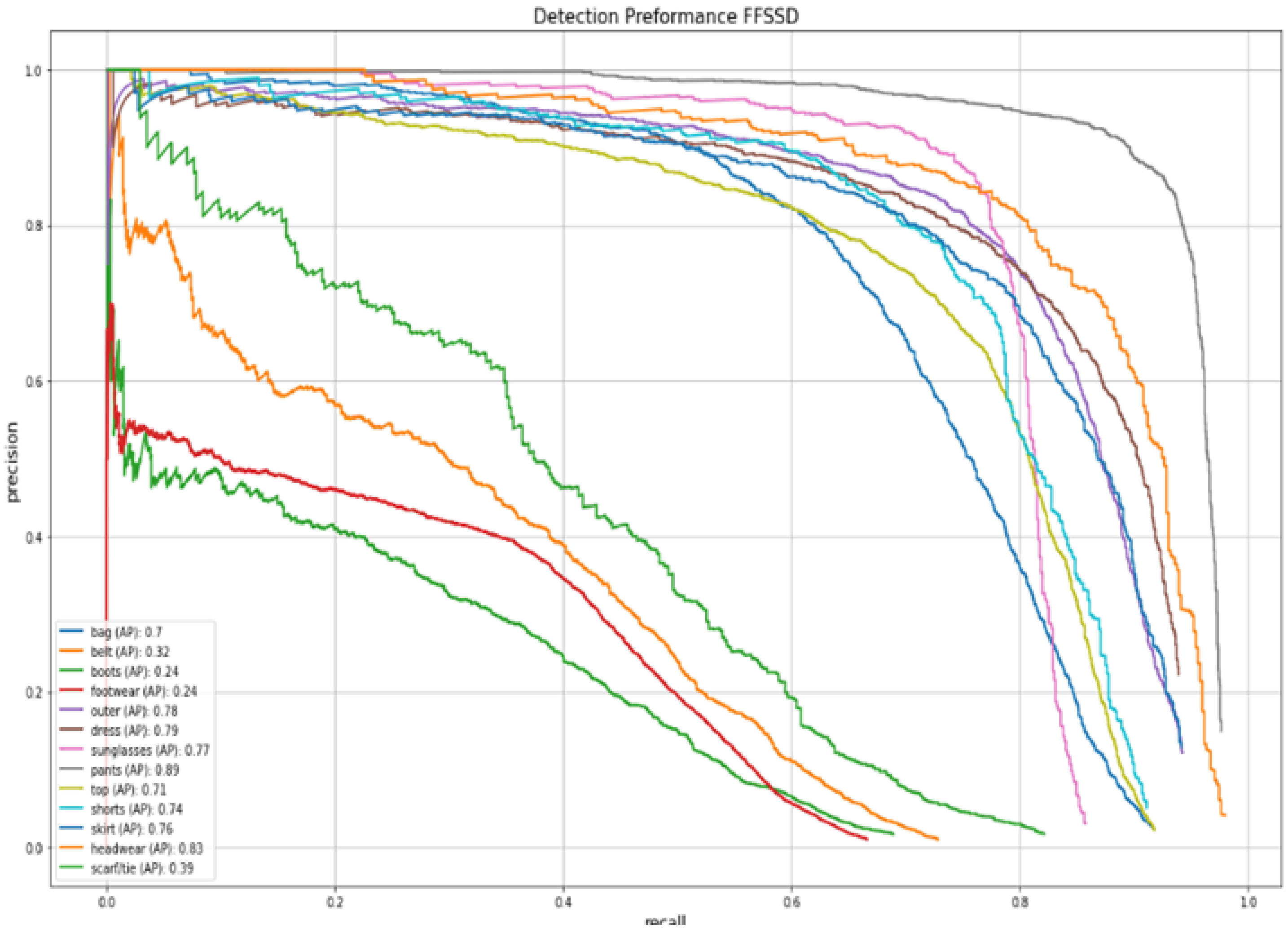


With the FFSSD the loss graph of the network tends to overfitting earlier than the SSD

UNIVPM EXPERIMENTS

Test Evaluation for FFSSD

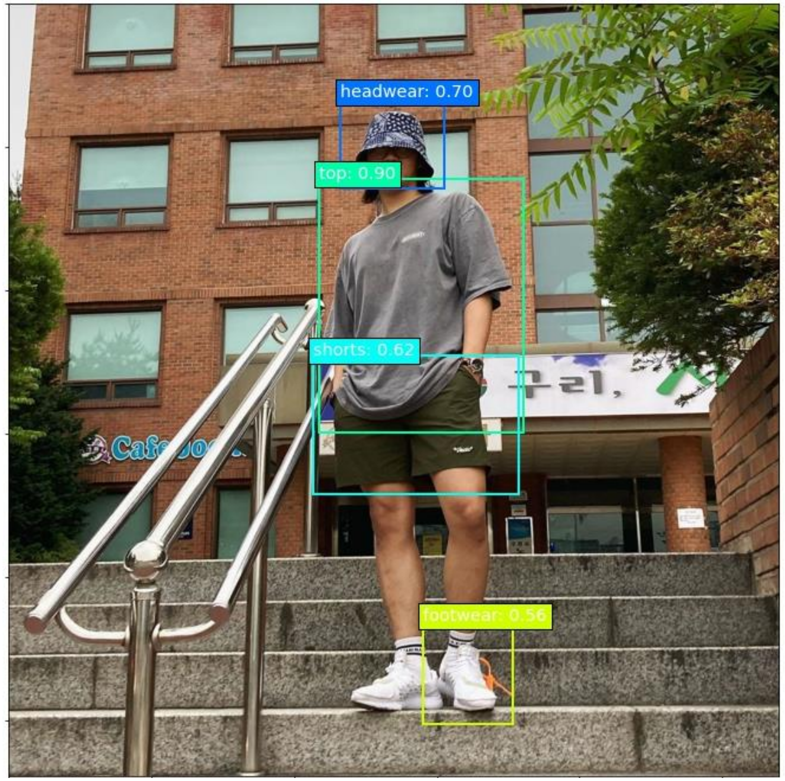
bag	AP	0.697
belt	AP	0.318
boots	AP	0.243
footwear	AP	0.243
outer	AP	0.782
dress	AP	0.788
sunglasses	AP	0.77
pants	AP	0.888
top	AP	0.707
shorts	AP	0.738
skirt	AP	0.765
headwear	AP	0.828
scarf/tie	AP	0.391
All categories	mAP	0.627

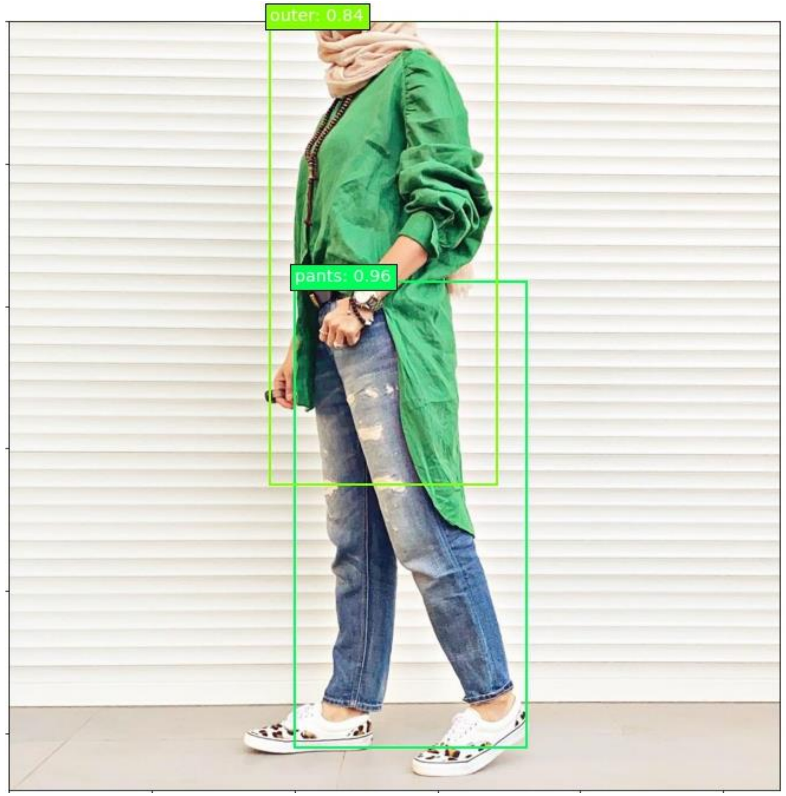


UNIVPM EXPERIMENTS

Small Object Detection

bag	AP	0.549
belt	AP	0.15
boots	AP	0.185
footwear	AP	0.14
outer	AP	0.769
dress	AP	0.763
sunglasses	AP	0.078
pants	AP	0.874
top	AP	0.639
shorts	AP	0.607
skirt	AP	0.697
headwear	AP	0.478
scarf/tie	AP	0.295
	mAP	0.479



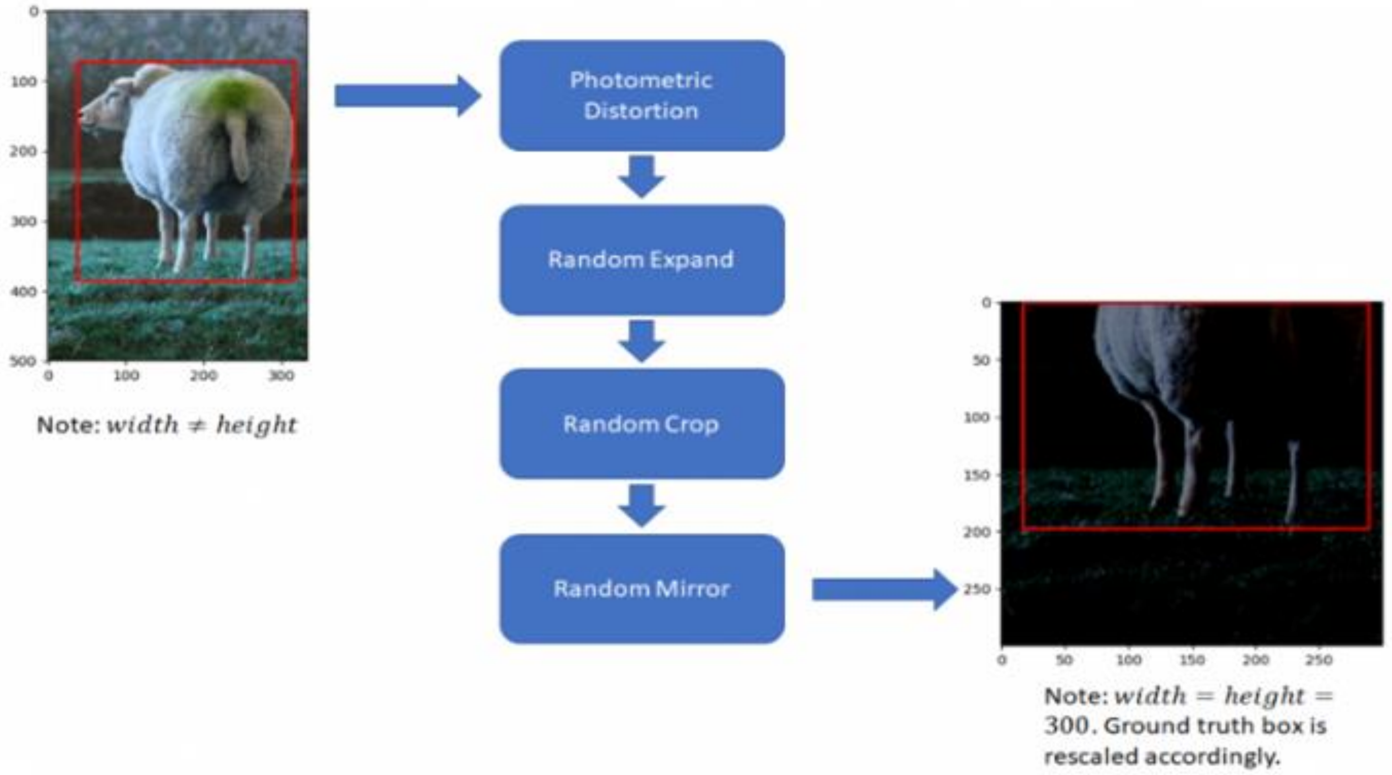


UNIVPM EXPERIMENTS

Data augmentation

The following data augmentation steps are used and are applied in the order listed:

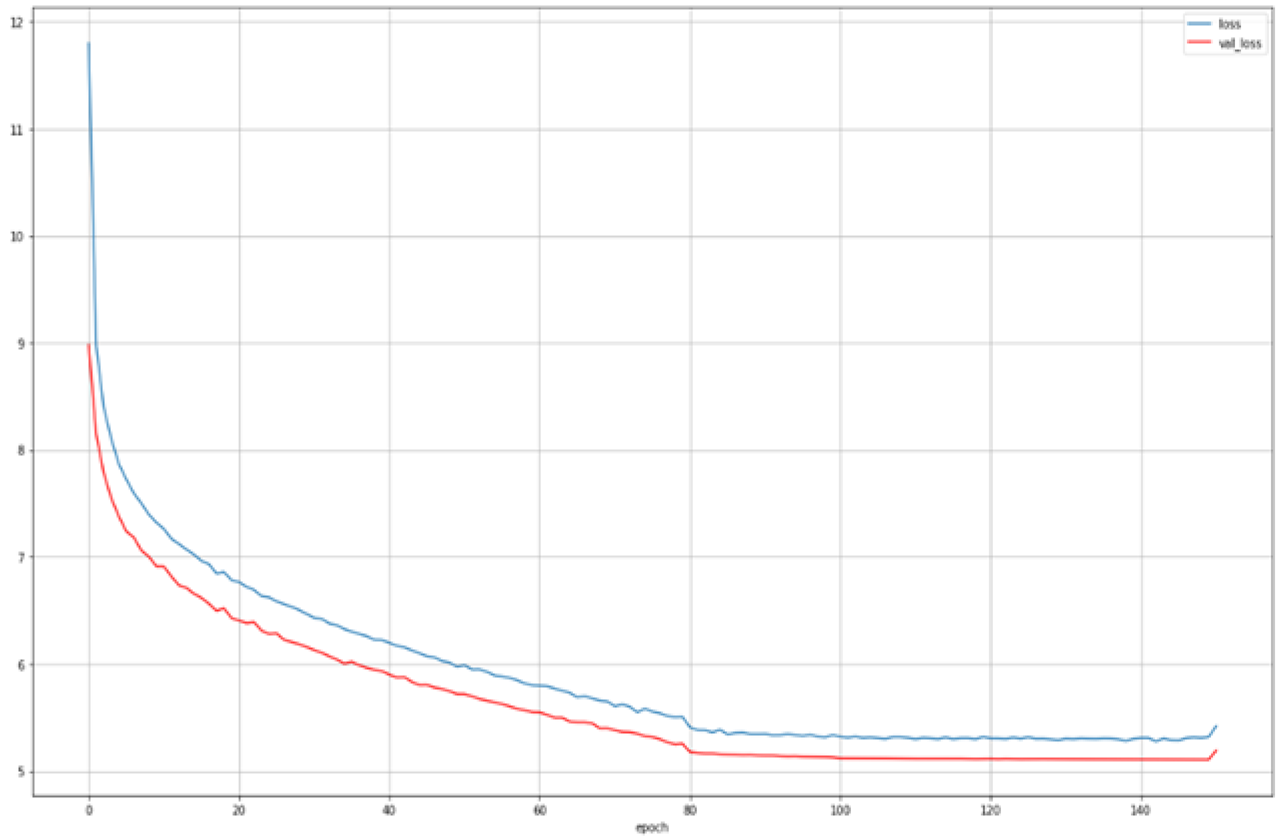
- Photometric Distortions
  - Random Brightness
  - Random Contrast, Hue, Saturation
  - RandomLightingNoise
- Geometric Distortions
  - ExpandImage
  - RandomCrop
  - Mirror



UNIVPM EXPERIMENTS

Data augmentation

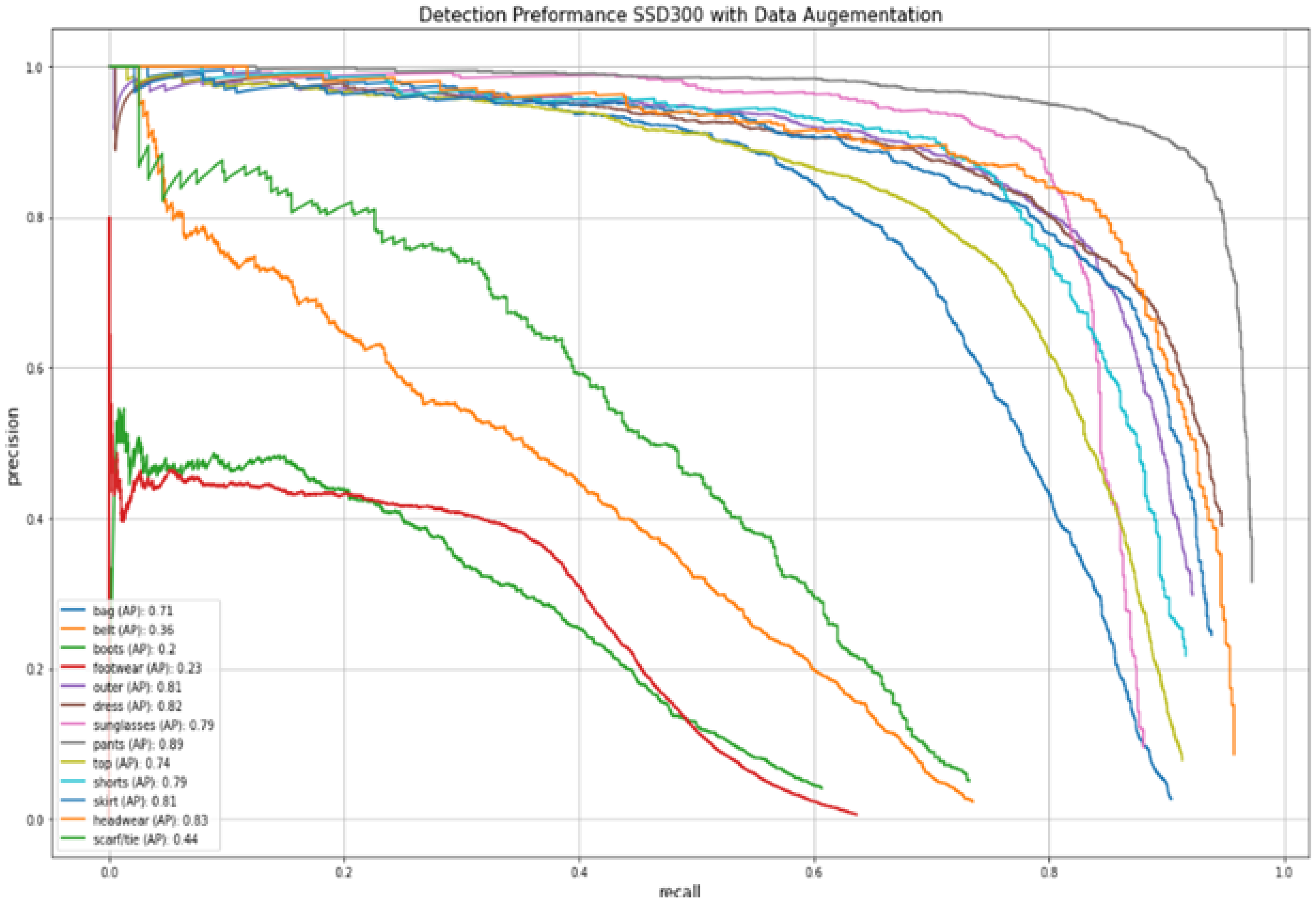
With the date augmentation applied to the SSD 300 the network needs more than 120 epoch, in fact as we can see from the graph of the loss it continues to improve.
So we ended the train at epoch 150.



UNIVPM EXPERIMENTS

Data augmentation

bag	AP	0.712
belt	AP	0.361
boots	AP	0.204
footwear	AP	0.231
outer	AP	0.809
dress	AP	0.821
sunglasses	AP	0.792
pants	AP	0.889
top	AP	0.745
shorts	AP	0.795
skirt	AP	0.809
headwear	AP	0.829
scarf/tie	AP	0.443
All categories	mAP	0.649

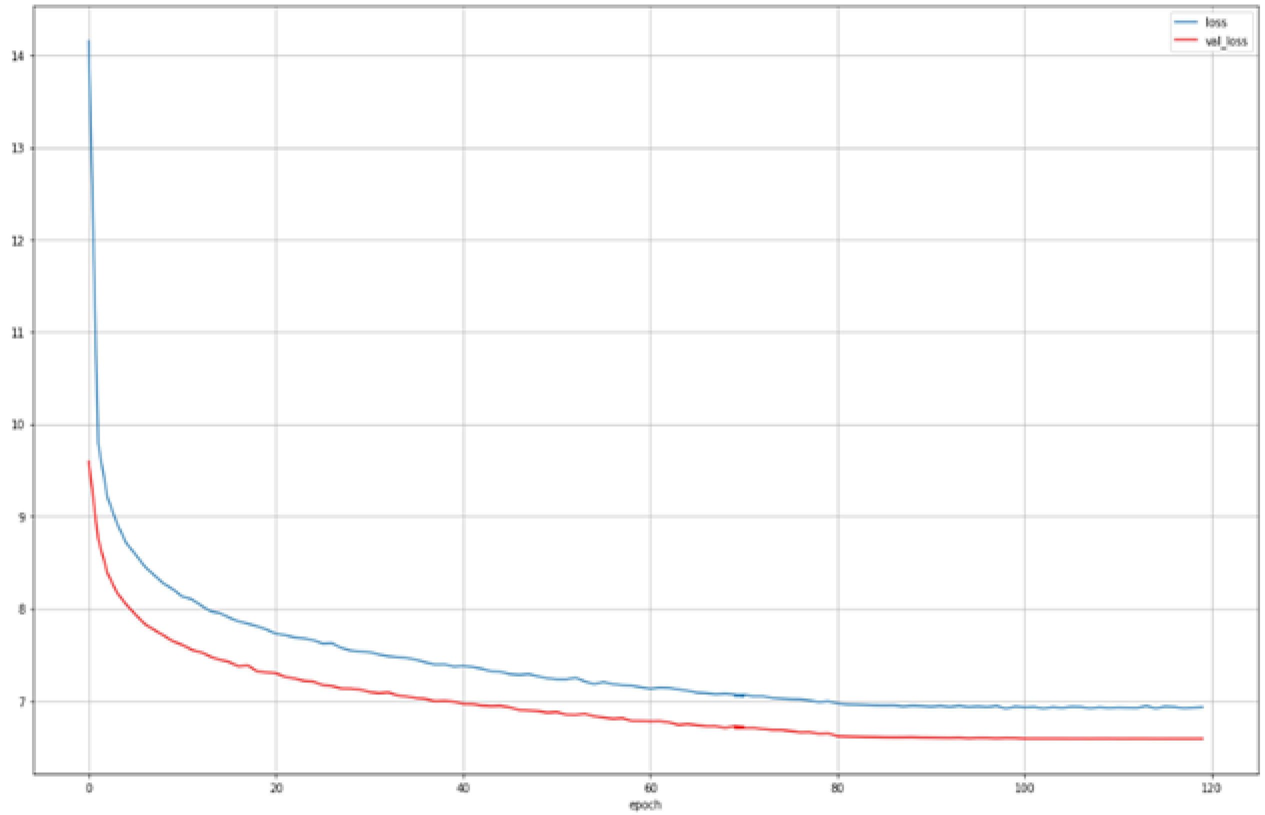


UNIVPM EXPERIMENTS

Data augmentation

FFSSD with this parameter:

- Batch size 64
- 120 epoch
- 256 iteration per epoch

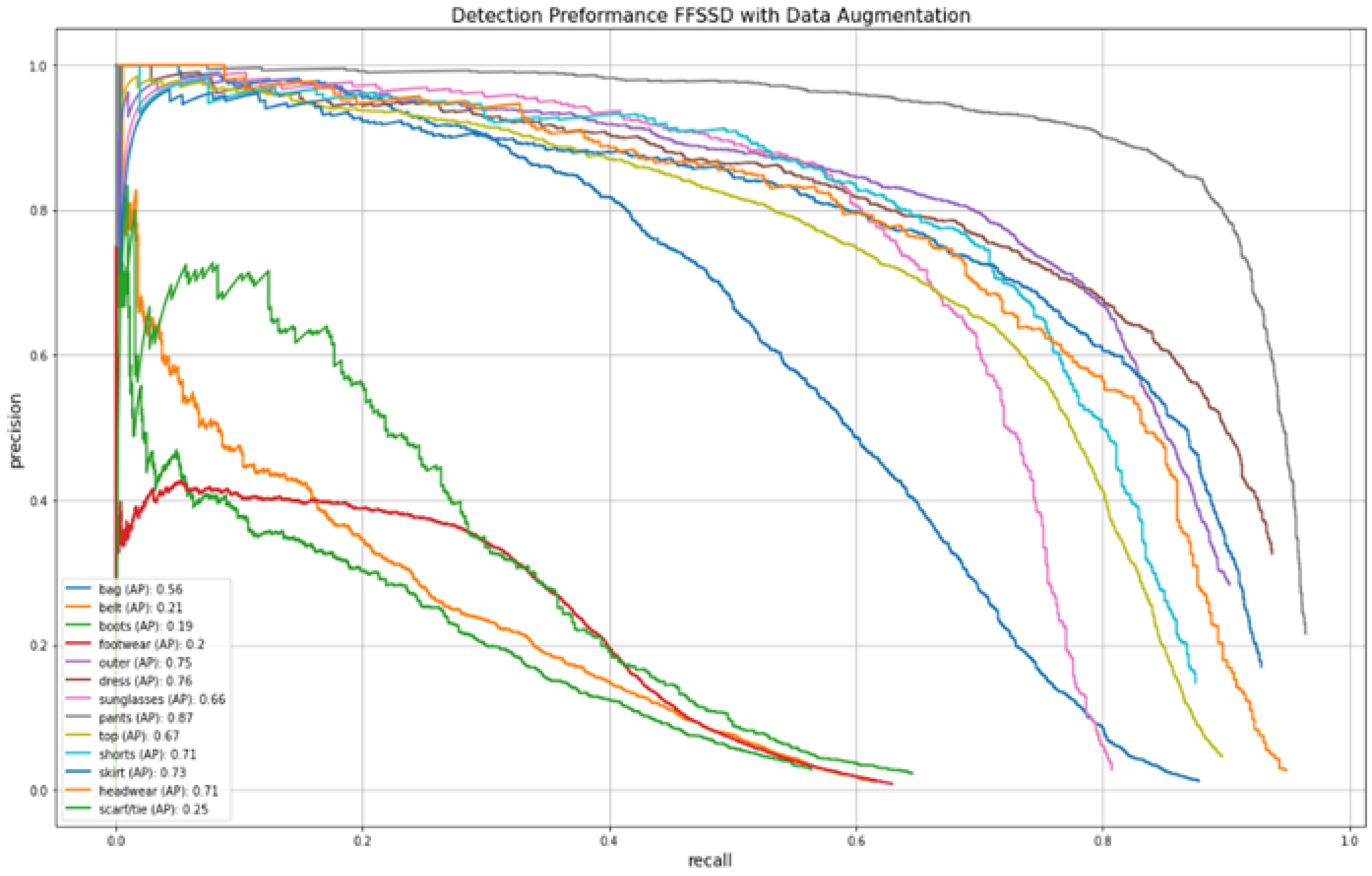


UNIVPM EXPERIMENTS

Data augmentation

The evaluation of FFSSD with data augmentation, shows us that there is a worsening of the results compared to the FFSSD without data augmentation.

bag	AP	0.562
belt	AP	0.209
boots	AP	0.188
footwear	AP	0.197
outer	AP	0.753
dress	AP	0.761
sunglasses	AP	0.657
pants	AP	0.866
top	AP	0.665
shorts	AP	0.708
skirt	AP	0.726
headwear	AP	0.715
scarf/tie	AP	0.252
All categories	mAP	0.558



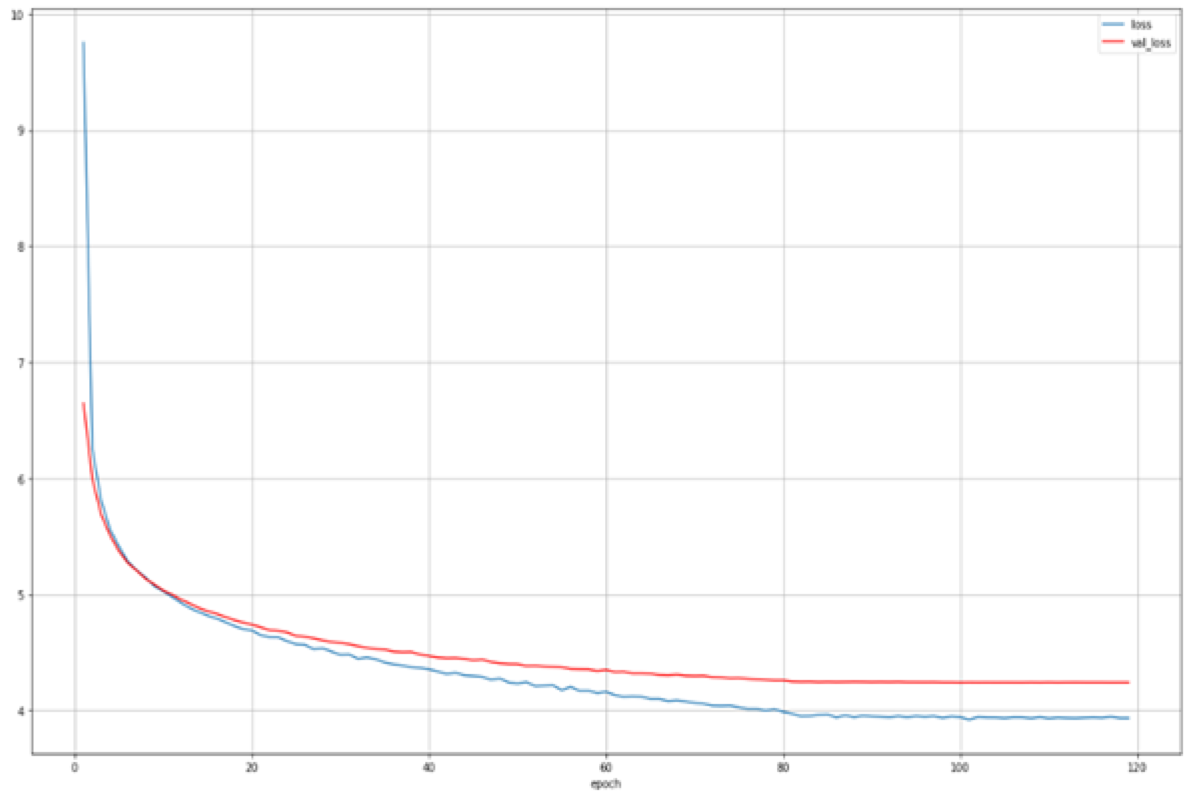
UNIVPM EXPERIMENTS

	SSD300 (AP)	FSSD (AP)	SSD300 D_A(AP)	FFSSD D_A(AP)
Bag	0.674	0.697	**0.712**	0.562
Belt	0.285	0.318	**0.361**	0.209
Boots	**0.26**	0.243	0.204	0.188
Footwars	**0.27**	0.243	0.231	0.197
Outer	0.785	0.782	**0.809**	0.753
Dress	0.784	0.788	**0.821**	0.761
Sunglasses	0.79	0.77	**0.792**	0.657
Pants	0.885	0.888	**0.889**	0.866
Top	0.705	0.707	**0.745**	0.665
Shorts	0.74	0.738	**0.795**	0.708
Skirt	0.754	0.765	**0.809**	0.726
Headware	0.809	0.828	**0.829**	0.715
Scarf/Tie	0.387	0.391	**0.443**	0.252
mAP	0.625	0.627	**0.649**	0.558

UNIVPM EXPERIMENTS

Fine tuning

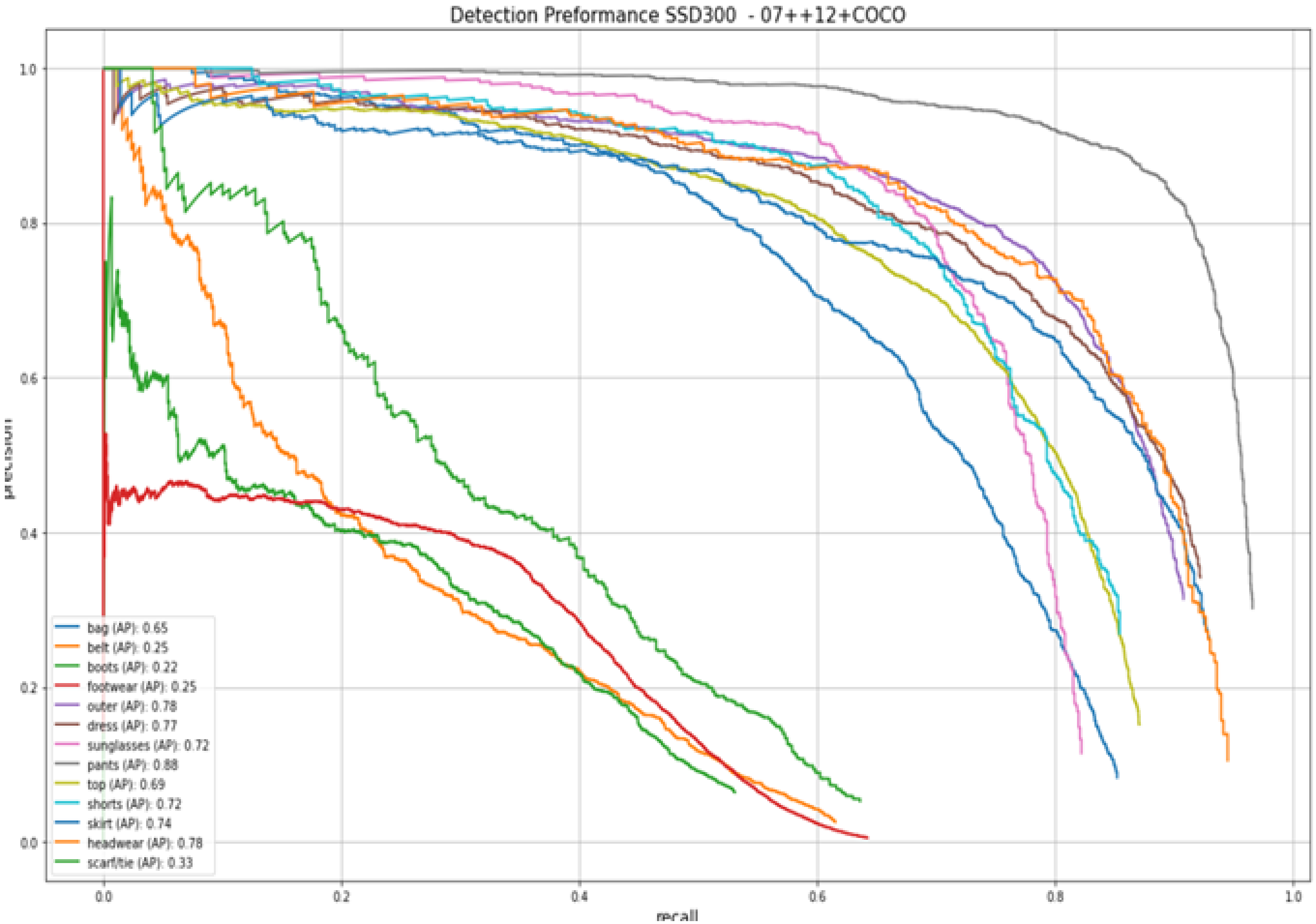
We implement the fine tuning of SSD300 with the original models of PASCAL VOC models (07++12+COCO).



UNIVPM EXPERIMENTS

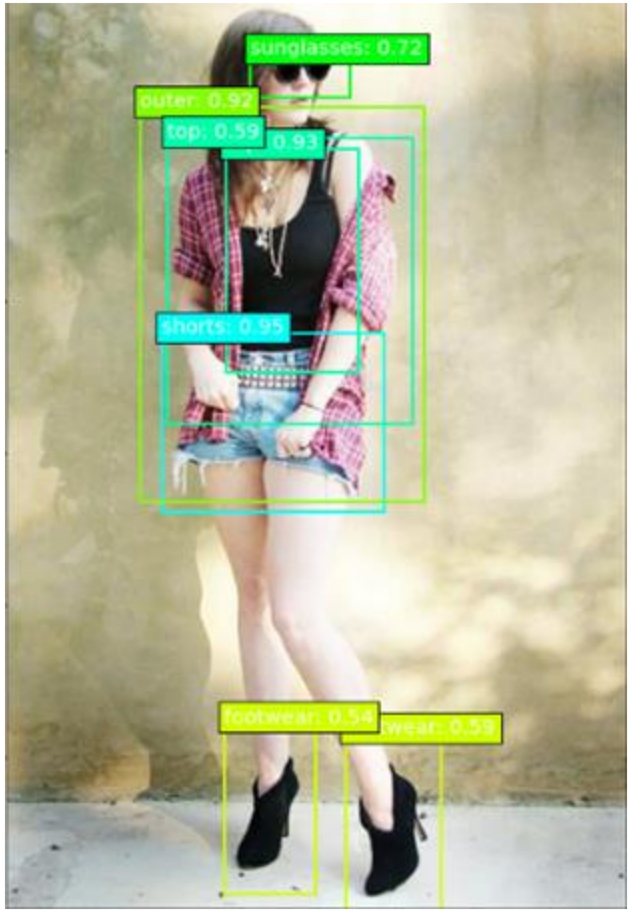
Fine tuning

bag	AP	0.649
belt	AP	0.253
boots	AP	0.217
footwear	AP	0.246
outer	AP	0.777
dress	AP	0.771
sunglasses	AP	0.719
pants	AP	0.877
top	AP	0.693
shorts	AP	0.718
skirt	AP	0.744
headwear	AP	0.782
scarf/tie	AP	0.332
All categories	mAP	0.598



UNIVPM EXPERIMENTS

Visual Evaluation on the test set



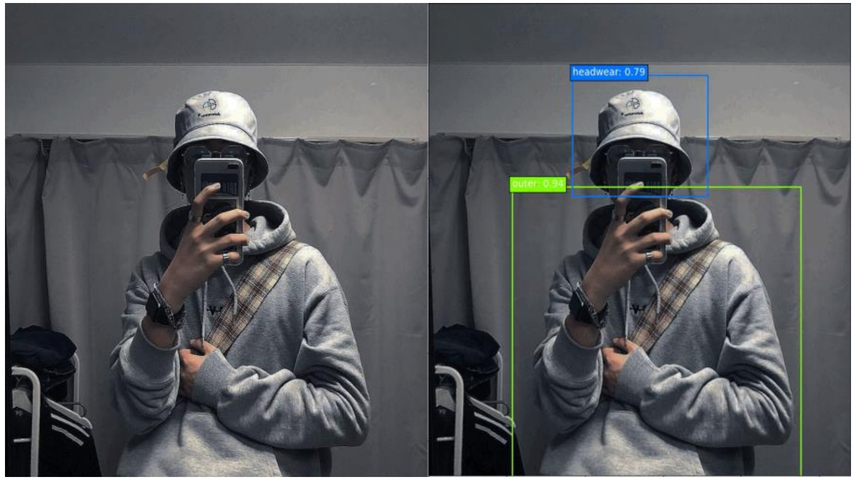
Here we use the test set images to evaluate the results of the detection…

BUT THIS IS SIMPLE…

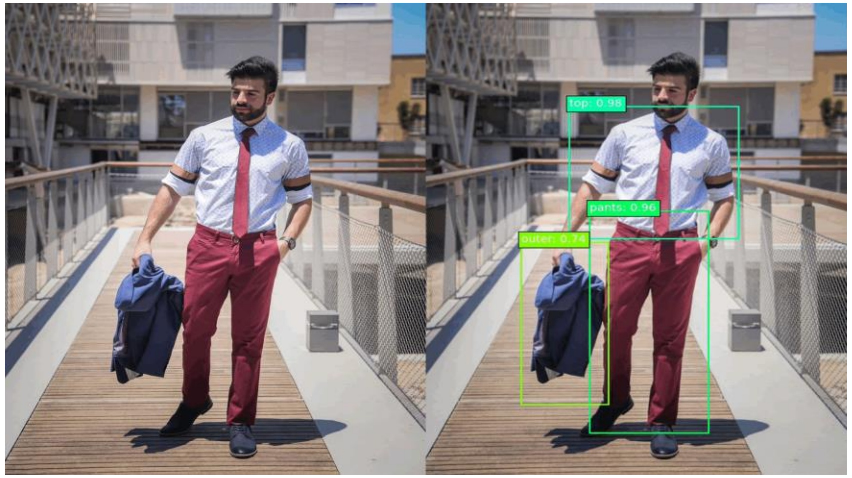
UNIVPM EXPERIMENTS

Visual Evaluation on Instagram Images

We download data from Instagram using three #hashtag:



#outfit



#streetoutfit

#streetstyle

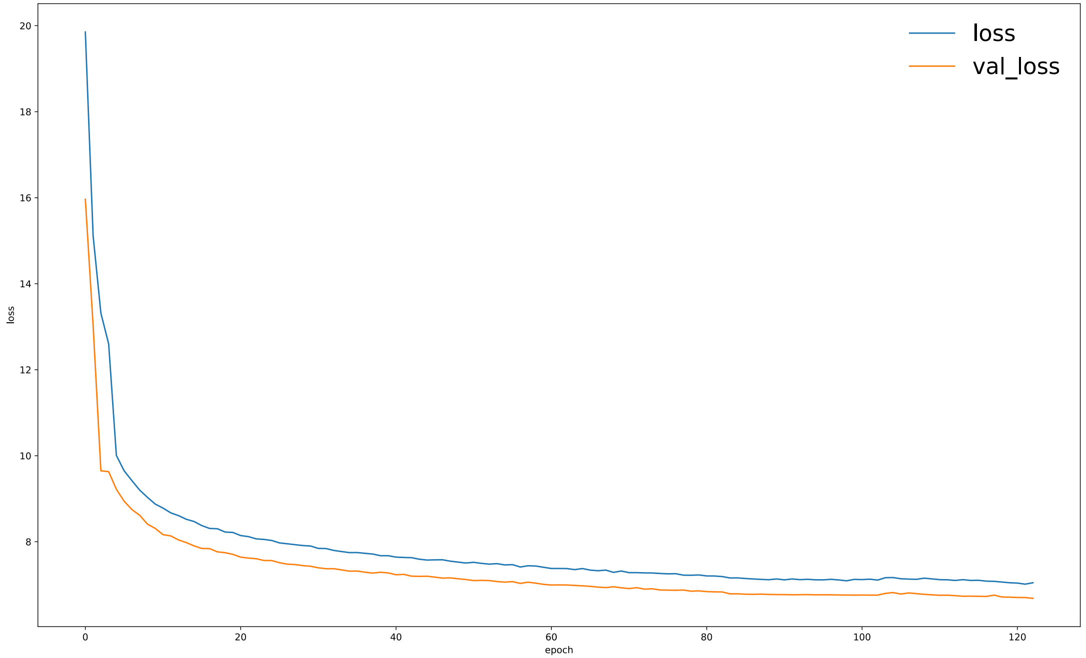
COMPUTATIONAL TIME COMPARISON

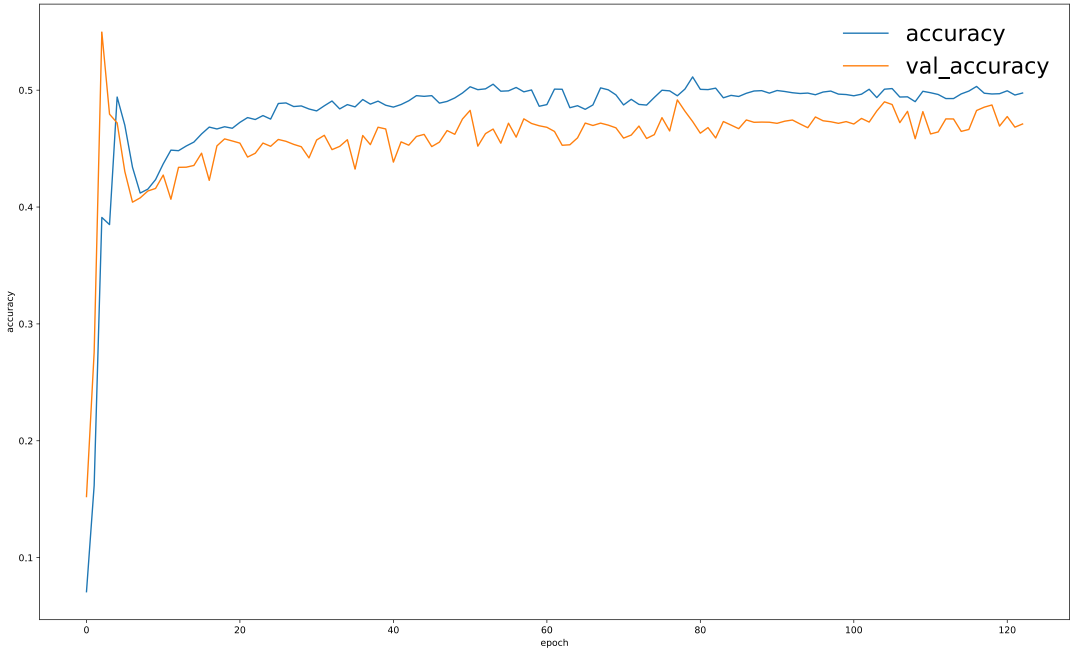
UNIVPM Architecture		ENEA	Google Colab
RTX 2080 TI (11GB GDDR6) 4352 CUDA Cores	Quadro P3200 (6GB GDDR5) 1792 CUDA Cores	Tesla K40m (12GB GDDR5) 2888	Tesla P100 (16GB HBM2) 3584 CUDA Cores
16 hours (120 epochs)	4 days (120 epochs) not finished yet	N/A*	24 hours (120 epochs)

* The framework version do not support this GPU. The Hardware CUDA Version is holdest (3.5 instead of 5.2 required)

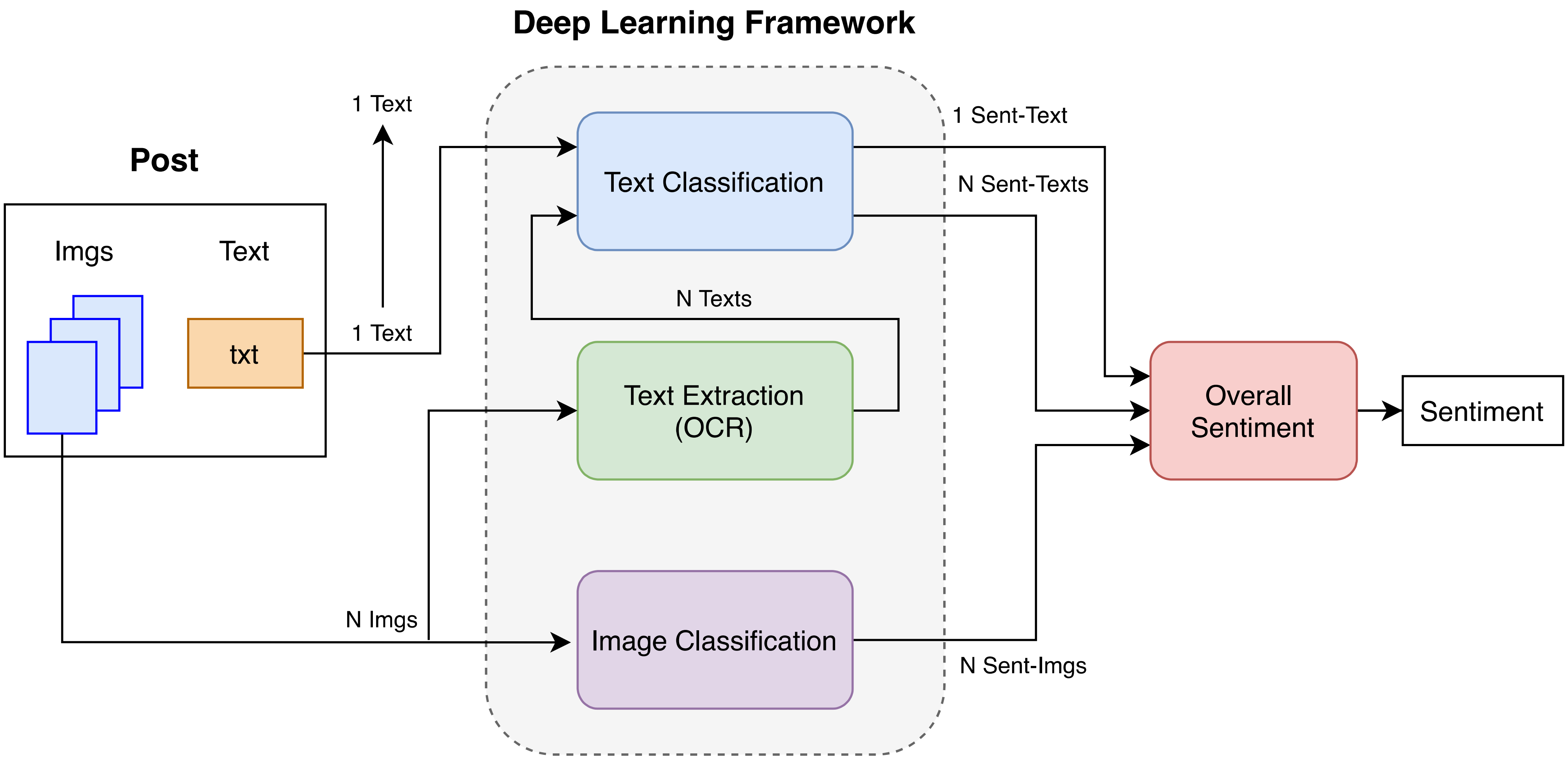
DL-Application

SSD – Training results – UNIVPM - RTX





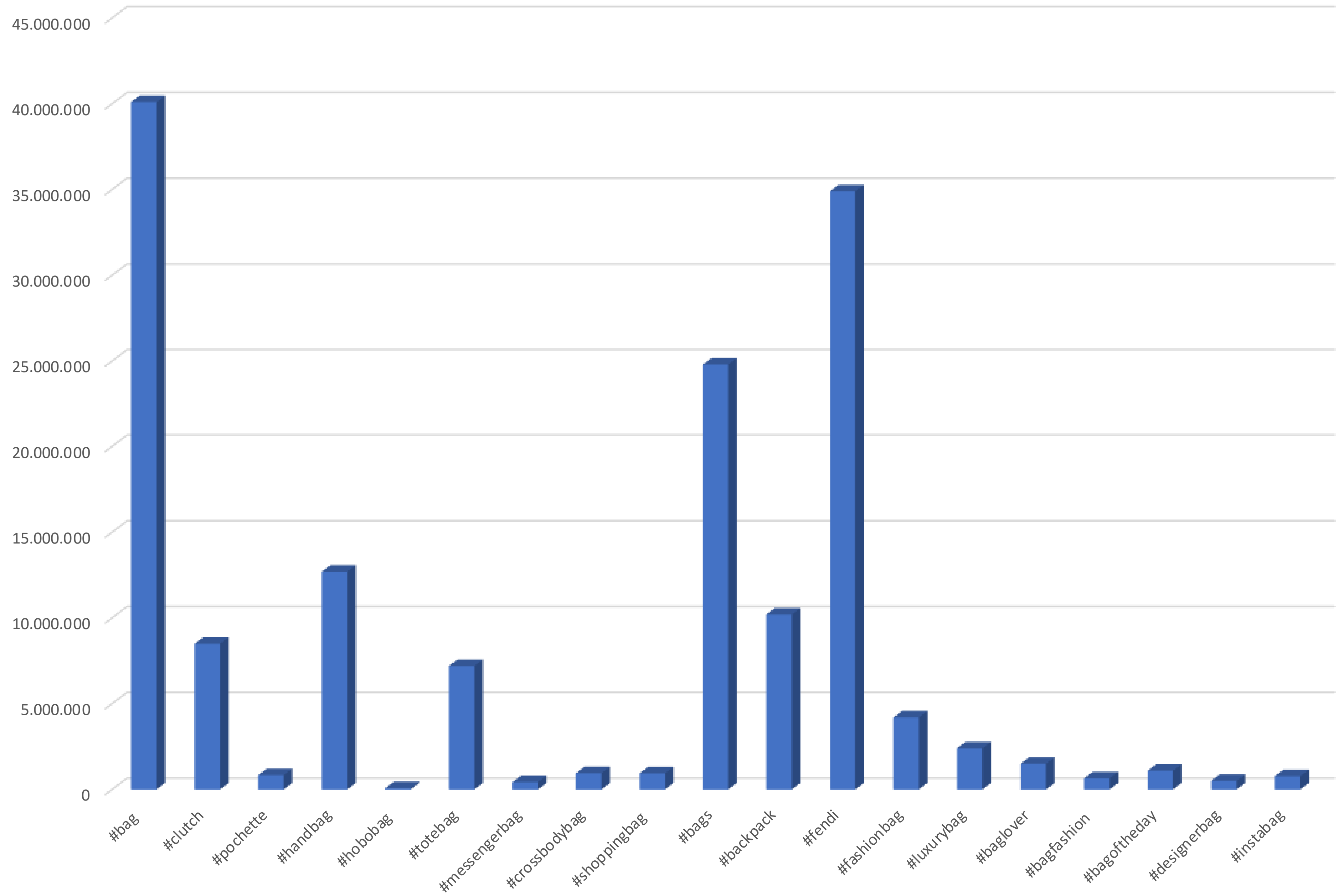
USE CASES – REAL EXAMPLES



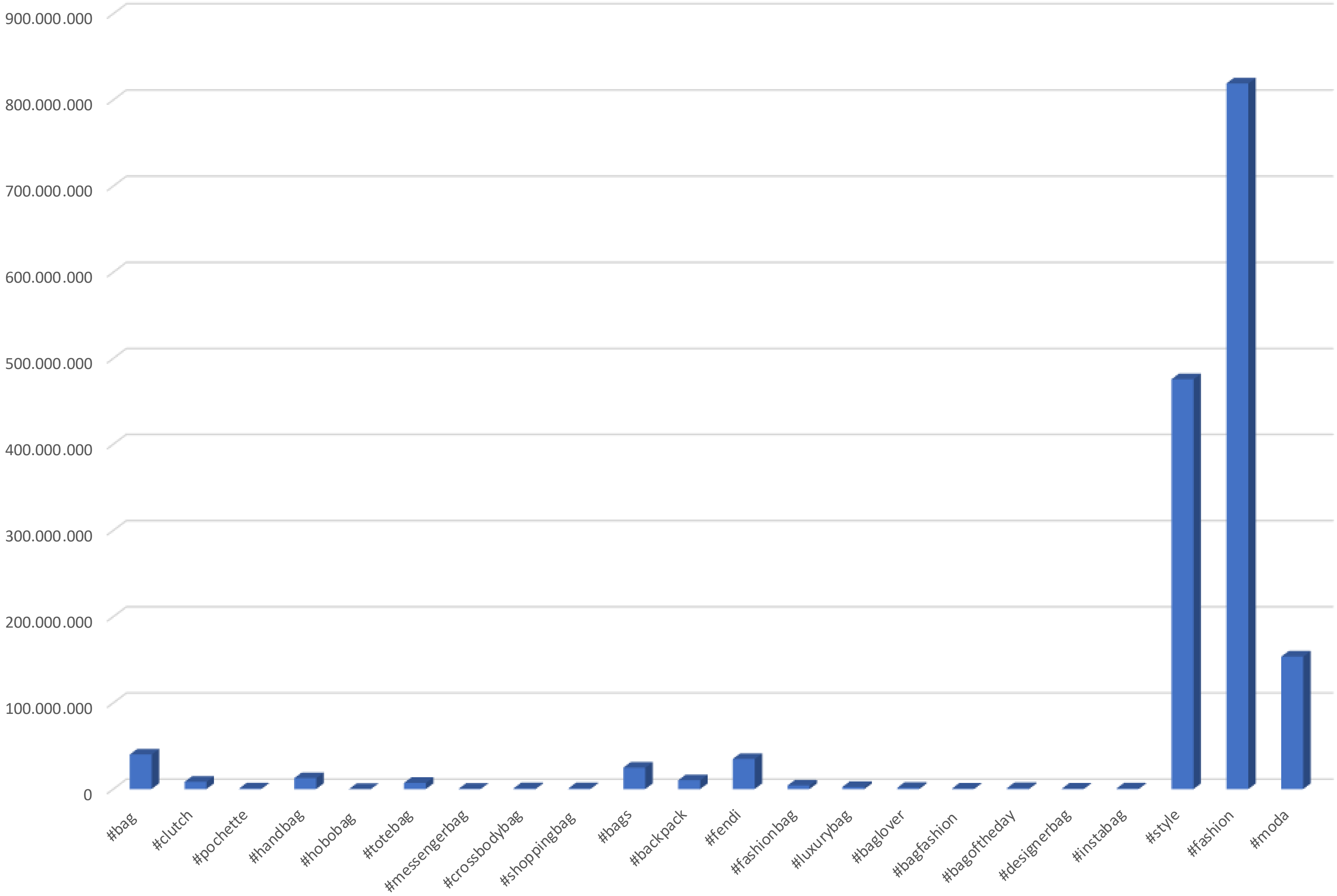
USE CASES – REAL EXAMPLES

- Copy of social activities for future investigations (Posts, comments, influencers, sentiment, media, geolocation)
- Identification of "activity" peaks (n post + m like)
- Time trend analysis (geographical origin, sentiment)
- Detection of negative sentiment peaks
- Inspection of Influencers due to high activity / low sentiment period
- Multi-tracker Influencer
- Geographical origin of the influencers cause of the high activity / low sentiment period
- Exclusion users and hashtag
- Average / non-average

DATA FROM IG-HASHTAG



DATA FROM IG-HASHTAG



DISCUSSIONS AND OPEN QUESTIONS

- Analysis and Executive Design
- Development of a tagging tool
- Adjustment of the learning systems present in UNIVPM laboratories and development of the complete learning system and proposal of a subset of results to the user for manual verification
- Management of the test phase, analysis of results, assistance and improvement of the approaches in progress